CHAPTER II

REVIEW OF THE LITERATURE

Introduction

The opportunity for evaluation practitioners to use microcomputer programs when conducting content analyses of responses to open-ended survey questions now exists. However, the best use of this opportunity requires a sound understanding of the conceptual and operational relationships between evaluation, content analysis, and microcomputers. This chapter is intended to promote such an understanding so that evaluation practitioners can better address applied content analysis problems used in their work. It is also used to provide the basis for the experimental design elements of this study. The following discussions are organized into four main topics.

The first topic focuses on key concepts for identifying the relationships between evaluation and content analysis. These concepts are discussed in terms of different types of information, actions, and standards of quality. Working definitions of evaluation and content analysis that use these concepts are then presented.

The second topic focuses on an overview of evaluation at two levels of detail. First, classification systems that have been used to describe alternative approaches to conducting evaluations are presented and consolidated into one framework. Second, each of the alternative approaches is presented in further detail. Content

analysis is placed into this framework as one type of evaluation approach.

Third, a general model for conducting an evaluation effort is presented. This model is based on the working definition of evaluation. As such, it is general enough to accommodate content analysis as part of an evaluation effort.

Fourth, an overview of content analysis focuses on its uses, tasks, and microcomputer implementations. In the first two discussions, several uses of content analysis are classified and then summarized. In the summary, those uses particularly relevant to evaluation are highlighted. Then, tasks crucial to obtaining high quality content analysis information are discussed. Finally, the ways microcomputers can help people perform many of these tasks are presented. These operational relationships provide the basis for designing focused studies on how microcomputers can be used to help people perform more reliable and valid content analyses used in evaluation efforts.

Key Concepts for Identifying the Relationships Between Evaluation and Content Analysis

Evaluation is often thought of as the process of describing and judging some object (e.g., Guba & Lincoln, 1981; Joint Committee, 1981; Stake, 1967; Worthen & Sanders, 1973), while content analysis is often thought of as the process of describing and making inferences about some object (e.g., Holsti, 1969; Osgood, 1959; Stone, Dunphy, Smith & Ogilvie, 1966). Even though most authors use verb

forms of these concepts to represent actions, they are better thought of in their noun forms for this study--descriptions, judgments, and inferences--as types of information. As such, both of these enterprises have in common the process of developing a body of information about some object. Nevertheless, the basic actions people perform in order to develop evaluation or content analysis information also turn out to be quite similar (e.g., Krippendorff, 1980; Stufflebeam et al., 1971). The main difference between evaluation and content analysis is based on the underlying contrasts used to partition the information. These different underlying contrasts lead to different connotations for common terms, and different standards for judging the quality of practice. Because these are important issues, this section is used to (a) identify the key components of evaluation and content analysis information in terms of their underlying relationships, (b) identify basic actions used to develop evaluation and content analysis information, (c) discuss different standards of quality that have emerged for judging the information and related processes, and (d) present working definitions of evaluation and content analysis.

Key Components of Evaluation and Content Analysis Information

Even though evaluation can be used to develop descriptions and judgments about an object while content analysis can be used to develop descriptions and inferences about it, this partitioning is not as clear cut as it first appears. The term, descriptions, does not have quite the same meaning to evaluation theorists as it does to

content analysis theorists. In addition, the terms, judgments and inferences, are more similar in meaning to evaluation and content analysis theorists than one might first expect. Fortunately, the conceptual similarities and differences between the various types of information can be clarified by examining the underlying conceptual contrasts and identifying other terms that more clearly reflect these contrasts. This is accomplished by identifying the applicable underlying contrasts, combining them to clarify their relationships, and redefining the key types of evaluation and content analysis information based on these relationships. The main benefit from this examination is a more refined perspective on the fundamental characteristics of both evaluation and content analysis information.

Evaluation Contrasts

Webster's dictionary (Gove, 1971, p. 786) defines the term, evaluate, as "to examine and judge concerning the worth, quality, significance, amount, degree or condition of." The American Heritage dictionary (Morris, 1969, p. 453) defines the term, evaluate, as "1. To ascertain or fix the value or worth of. 2. To examine and judge; appraise." These definitions support two basic premises for this study. First, evaluation is values-based. This premise is important because it implies all true evaluation efforts must be values-based and any efforts that are not values-based are not evaluation efforts. They are something else. This premise is also consistent with Scriven's contention that evaluation "goals always include the estimation of merit, worth, [or] value" (1967, p. 42).

The second premise is that evaluation involves two fundamental information components. For now, the terms descriptions and judgments will be used to represent these components. Many definitions of evaluation include these two components, although at varying levels of specificity. The dictionaries refer to these components as examination and judgment. Stake (1967) claims, "Both description and judgment are essential--in fact, they are the two basic acts of evaluation" (p. 109). Worthen and Sanders (1973) state, "Evaluation is the determination of the worth of a thing. It includes obtaining information for use in judging the worth of [some object]" (p. 19). Guba and Lincoln (1981) "define evaluation as a process for describing an evaluand and judging its merit or worth" (p. 35). Finally, the Joint Committee on Standards for Evaluations of Educational Programs, Projects, and Materials (Joint Committee, 1981) defines evaluation as "the systematic investigation of the worth or merit of some object" (p. 12).

Thus, the typical or surface contrast made between two fundamental components of evaluation information is that of descriptions vs. judgments. This contrast is based on an underlying contrast that has its roots in philosophy. Epistemology is the philosophy of knowledge. It addresses questions about how we can discover and know the truth or facts about some object. Ethics is the philosophy of values. From ethics we determine what is good or bad, what is right or wrong, and what we ought to do in a given situation--rules of conduct. Judgments are ultimately based on ethical principles--value statements.

Content Analysis Contrasts

Berelson (1952) provides the classic definition of content analysis. He states, "Content analysis is a research technique for the objective, systematic, and quantitative description of the manifest content of communication" (p. 18). This descriptive orientation represents the classical content analysis approach. However, many theorists consider it to be too narrow.

A broader perspective of content analysis goes beyond description to also include making inferences. For example, Osgood (1959) defines "content analysis as a procedure whereby one makes inferences about sources and receivers from evidence in the messages they exchange" (p. 36). A similar definition was developed jointly by Stone and Holsti, although they published it separately. They contend, "Content analysis is any research technique for making inferences by systematically and objectively identifying specified characteristics within text" (Holsti, 1969, p. 14; Stone et al., 1966, p. 5).

Thus, content analysis is often thought of as involving the process of describing some object and making inferences related to it. The surface contrast is between descriptions vs. inferences. The underlying contrast is based on the means through which information about some object is acquired. Descriptions are based on sensory input from the outside world. Regardless of the actual senses involved (sight, hearing, touch, taste, or smell) or the actual perceiver (human or otherwise), all sensory input can be called observations. Inferences derive from applying the rules of logic to

a set of statements. These statements usually include descriptions based on observations, and other statements--conclusions--based on theoretical principles. The theories and their principles can be grounded in either epistemology or ethics.

The Contrasts Combined

Table 1 is used to summarize the surface contrasts and underlying contrasts for evaluation and content analysis information used in this study. The surface contrast is descriptions vs. judgments for evaluation, and descriptions vs. inferences for content analysis. The underlying contrast is knowledge vs. value statements for evaluation, and observations vs. logic for content analysis.

When the underlying contrasts are crossed in a two-by-two table, four types of information can be identified. This is represented in Table 2. The four types of information include observational knowledge, logical knowledge, observational value statements, and logical value statements.

Observational knowledge is typified by what we commonly call "the facts"--what we can learn simply by observing something. "The union has 197 members," represents a statement of observational knowledge.

Logical knowledge is derived from applying rules of logic to available information and knowledge-seeking, theoretical principles. "The experimental treatment effect was significantly greater than the control treatment effect," represents a statement of logical knowledge.

Table 1

Surface and Underlying Contrasts for Evaluation and Content Analysis Information

Contrast	Evaluation	Content analysis
Surface	Descriptions vs. Judgments	Descriptions vs. Inferences
Underlying	Knowledge vs. Value statements	Observations vs. Logic

Table 2

Four Types of Information Derived From Underlying Evaluation and Content Analysis Contrasts

Content analysis contrast	Evaluation contrast	
	Knowledge vs.	Value statements
Observations vs. Logic	Observational knowledge	Observational value statements
	Logical knowledge	Logical value statements

Observational value statements of good/bad, right/wrong, or how we should act in a particular situation, are based on the assumption we can observe the value of something or someone in the same way we can observe many of its other attributes, such as color or language spoken. Such an assumption is based on the philosophical doctrine called ethical naturalism (Harrison, 1967). "According to ethical naturalism, moral judgments just state a special subclass of facts about the natural world" (Vol. 3, p. 69). This doctrine has been rejected by G. E. Moore. Harrison, (1967) represents Moore's position as follows:

> Moore contended that goodness was a unique, unanalyzable, nonnatural property (as opposed to natural properties, such as yellowness or anger, that are perceived through the senses or through introspection). Therefore, any attempt to define goodness in terms of any natural property must be a mistake that is one form of what he called the "naturalistic fallacy." (Vol. 3, p. 69)

Thus, observational value statements are examples of the naturalistic fallacy. Because of Moore's criticism and the availability of a commonly accepted alternative to be discussed next, observational value statements are not included as an acceptable type of information for this study.

Logical value statements are derived from applying rules of logic to available information and ethical principles. The naturalistic fallacy is not at issue in this situation because no contention is made that these statements represent factual attributes about the object in question. Instead, the statements derive their value components from the specific ethical principles involved.

Evaluators typically distinguish between two types of value statements (e.g., Guba & Lincoln, 1981; Joint Committee, 1981; Scriven, 1967) most often referred to as merit and worth. Guba and Lincoln (1981) use the term, merit, to mean "intrinsic, context-free value" (p. 39). They further state an entity has merit if it has "value of its own, implicit, inherent, independent of any possible applications" (p. 39). When an entity has value within some context of use or application, they use the term, worth. They define it to mean "extrinsic or context-determined value" (p. 40). They acknowledge their terms, merit and worth, are types of value and they are analogous to Scriven's (1978) terms, merit and value; but they claim the use of their terms avoids "the redundancy and confusion that result when one of the subtypes is called by the same name as the more general type" (p. 40). They also acknowledge Scriven's (1967) notions of intrinsic and payoff evaluation and Tyler's (1949) concern for internal checkpoints and desired outcomes allude to the distinctions they make between merit and worth. However, they contend they "have addressed the issue in a more systematic way than Tyler and other earlier writers" (p. 40).

The concept of merit sounds suspiciously like observational value statements, but this need not be the case. If merit is taken to mean the factual value component of some object, then it does represent the naturalistic fallacy and it is not an acceptable type of information for this study. On the other hand, if merit is taken to mean value implicitly <u>generalized</u> to become free of any specific context, then the naturalistic fallacy is avoided. From this

perspective, merit is not a factual attribute of an object but a generalized depiction of value for that object in relation to a class of contexts. For example, to say a university professor with several refereed publications in his or her field has merit, should not be taken to mean publications are a value-attribute of university professors. Instead, it should mean, generally speaking, professors with many refereed publications are of value (i.e., competent and productive). In this way, the generalized value statement can only be derived by combining information about the object and a class of contexts with ethical principles by using rules of logic.

Because this process is often performed implicitly, it can take on the appearance of an observational value statement. On closer examination, however, the ethical perspectives necessary to make determinations of merit can usually be extracted. To avoid the controversy associated with the naturalistic fallacy, the information, ethical principles, and logical transformations used to make determinations of merit should be explicitly stated.

The concept of worth represents context-specific value statements. As such, the naturalistic fallacy is not at issue because the value statements are clearly dependent on variable situations that include different and often conflicting ethical principles. This makes it impossible for them to be inherent attributes of an object.

The Key Components and Their Underlying Relationships

It is now possible to reconstruct the components of evaluation and content analysis information that distinguish them from each

other in terms of the underlying conceptual relationships involved. This reconstruction is represented in Figure 1. The terms in boxes represent key components, while the terms spanned by arrows represent underlying concepts. The concept of high quality information has been added to represent the implicit, common feature of all the components and other concepts. Standards of quality for each field will be the focus of a later discussion.

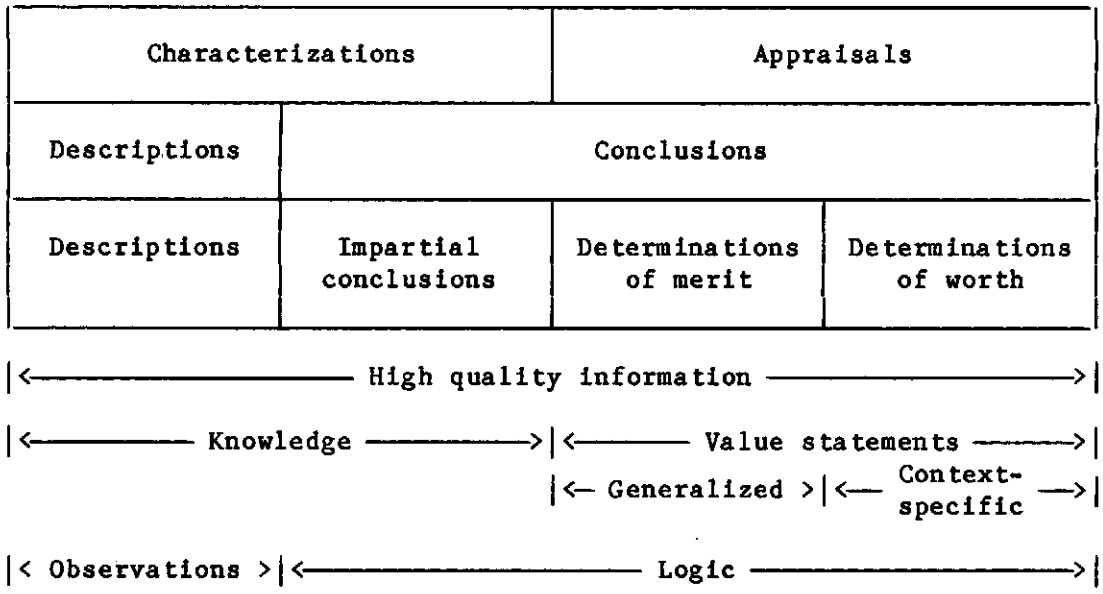


Figure 1. The Key Components of Evaluation and Content Analysis Information in Terms of Their Underlying Relationships

All concepts directly under a component help define the nature of that component. The terms for the components have been selected to represent the spirit of the applicable underlying concepts and complement the terms for the other components on all three levels. Also, the underlying concepts are arranged so that observations and

value statements do not overlap. This precludes any components from representing the naturalistic fallacy--considering value to be an inherent characteristic of an object.

The first level represents the two basic components of evaluation information. For the remainder of this study they will be called characterizations and appraisals. Characterizations are based on knowledge while appraisals are based on value statements.

The second level represents the two basic components of content analysis information. They will be called descriptions and conclusions. Descriptions are based on observations while conclusions are based on logic.

The third level represents four important subcomponents that are parts of both evaluation and content analysis information. These subcomponents include descriptions, impartial conclusions, determinations of merit, and determinations of worth. They are derived by simultaneously considering all applicable underlying concepts. Descriptions are based on observational knowledge. This subcomponent has the same name as a basic content analysis component because observational value statements have been excluded from consideration. Impartial conclusions are based on logical knowledge. Determinations of merit are based on logical, generalized value statements. Determinations of worth are based on logical, context-specific value statements.

All four subcomponents represent both evaluation and content analysis information, but they are grouped differently under the main components. For evaluation, characterizations are composed of

descriptions and impartial conclusions; while appraisals are composed of determinations of merit and determinations of worth. For content analysis, descriptions are not further subdivided; while conclusions are composed of impartial conclusions, determinations of merit, and determinations of worth.

Summary

The key components of evaluation and content analysis information, and the basis for their relationship have now been identified. The evaluation components, characterizations and appraisals, and the content analysis components, descriptions and conclusions, are based on different underlying contrasts. For evaluation, the key contrast is based on two fundamental branches of philosophy. Characterizations are based on epistemology, the philosophy of knowledge, while appraisals are based on ethics, the philosophy of determining good or bad, right or wrong, and rules of conduct. In modern usage, responses to ethical questions are often called statements of values (Frankenna, 1967). For content analysis, the key contrast is based on two modes of acquiring information about something. Descriptions are based on sensory input, observations. Conclusions are based on rules of logic that incorporate available information and theoretical principles intended to either acquire knowledge or derive value statements.

Four subcomponents of both evaluation and content analysis information have been identified by simultaneously considering the underlying contrasts of each field. One concept, observational value

statements, was excluded from further consideration because it represented the naturalistic fallacy--considering value to be a natural, inherent attribute of an object. It was replaced by further subdividing logical value statements into generalized and context-specific groups. All these components and underlying concepts are considered to represent high quality information. The actions through which this information is developed are the focus of the next section.

Actions of Evaluation and Content Analysis

The purpose of this section is to identify some very basic actions that can apply to different evaluation and content analysis approaches. These actions represent the operations evaluation and content analysis have in common. They also represent the action components of a general model to be discussed in a later section. Five sources, four from evaluation and one from content analysis literature, have been used to identify these actions.

Evaluation Actions

Stufflebeam et al. (1971) provide a good actions-oriented definition of evaluation. It is defined as "the (process) of [1] (delineating), [2] (obtaining), and [3] (providing) (useful) (information) for [4] (judging) (decision alternatives)" (p. 40). Each of the terms in parentheses are further defined by the authors.

The Joint Committee on Standards for Educational Evaluation (Joint Committee, 1981) has a set of thirty standards to be used for the action of evaluating evaluation. These standards are grouped in

a functional table of contents in terms of ten other actions: "[1] Administering Evaluation . . . [2] Analyzing Information . . . [3] Budgeting Evaluation . . . [4] Collecting Information . . . [5] Contracting Evaluation . . . [6] Deciding Whether to Evaluate . . . [7] Defining the Evaluation Problem . . . [8] Designing Evaluation . . . [9] Reporting Evaluation . . . [10] Staffing Evaluation" (pp. xvii-xx). Specific standards can apply to more than one action.

The Evaluation Research Society (ERS) has developed a set of fifty-five standards to be used for the action of evaluating evaluation (ERS Standards Committee, 1982). These standards are divided into six actions-oriented sections: "(1) Formulation and Negotiation, (2) Structure and Design, (3) Data Collection and Preparation, (4) Data Analysis and Interpretation, (5) Communication and Disclosure, and (6) Utilization" (p. 11). Each individual standard applies to only one of the above actions.

Brinkerhoff, Brethower, Hluchyj, and Nowakowski (1983) organize the chapters of their book, Program Evaluation: A Practitioner's Guide for Trainers and Educators, by seven evaluation functions. They call these functions: "1. Focusing the Evaluation . . . 2. Designing Evaluation . . . 3. Collecting Information . . . 4. Analyzing Information . . . 5. Reporting Information . . . 6. Managing Evaluation . . . 7. Evaluating Evaluation (Meta-Evaluation)" (p. v).

Content Analysis Actions

No formal classification system for content analysis actions like those for evaluation exists. However, one actions-oriented

classification system for content analysis activities comparable to those for evaluation was identified. Krippendorff (1980) uses this classification system as the basis for a practical guide to conducting a content analysis. He contends that "any content analysis involves three logically separate activities: [1] design, [2] execution, [3] report" (p. 169).

Consolidation and Summary

The above sources identify several actions that apply to both evaluation and content analysis efforts. However, many of the actions are too specific for what is needed here. When these actions are placed into more general groups, six basic actions of evaluation or content analysis efforts are suggested. The more specific actions are not intended to represent an exhaustive list of components that comprise the basic actions. The groupings have been used only to help identify those basic actions.

Due to size considerations, a comprehensive summary of evaluation tasks that can be subsumed under these six basic actions is not presented. However, the interested reader can find classifications of evaluation tasks from a number of other sources (e.g., Anderson, Soptick, Rogers, & Worthen, 1971; Schalock & Sell, 1972; Stufflebeam, 1973; Worthen & Sanders, 1984). A simple classification of content analysis tasks relevant to the experimental study is discussed in this chapter as part of the overview of content analysis.

Four of the actions focus on processing some kind of information while two focus on the effort itself. These six actions, their

relationships to the literature sources, and their main focus are summarized in Table 3.

First, the four actions of delineating, obtaining, providing, and applying information are suggested. (For evaluation, the two main types of information are characterizations and appraisals. For content analysis, the two main types of information are descriptions and conclusions.) Delineating information involves the general action of specifying what information is needed and how it will be acquired. Delineating encompasses more specialized actions identified in several sources. These actions were called contracting, deciding, defining, designing, formulating, focusing, negotiating, and structuring. Obtaining information involves the general action of acquiring it in its "raw" state and transforming it to a usable state. Obtaining is roughly synonymous with execution. It also encompasses more specialized actions. They were called analyzing, collecting, interpreting and preparing. Providing information involves the general action of delivering it to the appropriate audiences. Providing is similar to communicating, disclosing, and reporting. Applying information involves the general action of using it for intended or unintended purposes. Applying is an action at a level of specificity comparable with the three general actions above. It is roughly synonymous with utilizing. Judging decision alternatives is a specific example of applying (evaluation) information.

Second, the two actions of managing and evaluating an effort are suggested. These actions can be applied to both evaluation and content analysis efforts, even though Krippendorff does not directly

Table 3

Six Basic Actions for Conducting an Evaluation or Content Analysis Effort

Literature Source						
S	J	E	B	K	Action	Focus
1	5, 6, 7, 8	1, 2	1, 2	1	* Delineating	Information
2	2, 4	3, 4	3, 4	2	* Obtaining	
3	9	5	5	3	* Providing	
4		6			* Applying	
	1, 3, 10		6		* Managing	The effort
	All	All	7		* Evaluating	

Note. Numerals represent the actions identified in the text for the applicable source, where S = Stufflebeam et al., 1971, p. 40; J = Joint Committee, 1981, pp. xvii-xx; E = ERS Standards Committee, 1982, p. 11; B = Brinkerhoff, et al., 1983, p. v; K = Krippendorff, 1980, p. 169.

mention them. Managing an effort involves the general action of ensuring all required functions are performed appropriately. It is roughly synonymous with administering, and it encompasses the more specific actions of budgeting and staffing. Evaluating an effort involves the general action of characterizing and appraising it. All of the actions mentioned by the Joint Committee and the Evaluation Research Society were used in the context of evaluating an effort. Meta-evaluation is another term that can be used for evaluating evaluation efforts. The actions of interest for evaluating an effort were discussed in this section. Standards of quality for actually judging them are discussed in the next section.

Standards of Quality

Before evaluation or content analysis practice can be improved, some basis for determining what constitutes an improvement must exist. Standards of quality serve this role. Such standards are currently available for evaluation and content analysis practitioners but at different levels of formality. Evaluation practitioners have available to them published standards developed by professional and regulatory sources, while content analysis practitioners do not. Instead, they must rely on informal sources for indicators of quality. Some of these sources for each field are discussed next.

Evaluation Standards

Two sources of the most comprehensive standards for evaluation quality have already been presented. They are the Joint Committee on

Standards for Educational Evaluation (Joint Committee, 1981) and the Evaluation Research Society (ERS Standards Committee, 1982).

Joint Committee. The thirty Standards for Evaluations for Educational Programs, Projects, and Materials (Joint Committee, 1981)

> are presented in four groups that correspond to four main concerns about any evaluation--it utility, feasibility, propriety, and accuracy. Each standard is explained and clarified through a commentary which includes an overview of intent, guidelines for application, common pitfalls, caveats (or warnings against being overzealous in implementing the standard), and an illustration of the standard's application. (pp. 1-2)

Eight "Utility Standards are intended to ensure that an evaluation will serve the practical information needs of given audiences" (p. 19). Three "Feasibility Standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal" (p. 51). Eight "Propriety Standards are intended to ensure that an evaluation will be conducted ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results" (p. 63). Finally, eleven "Accuracy Standards are intended to ensure that an evaluation will reveal and convey technically adequate results about the features being studied that determine its merit or worth" (p. 97).

Evaluation Research Society. The fifty-five Evaluation Research Society Standards for Program Evaluation (ERS Standards Committee, 1982) are divided into six sections. The sections are listed in roughly sequential order for an evaluation effort. The individual standards are presented as "simple admonitory statements" (p. 11).

Twelve standards of Formulation and Negotiation are based on the assumption, "before an evaluation program or project is undertaken, the concerned parties should strive for a clear mutual understanding of what is to be done, how it is to be done and why, and for an appreciation of possible constraints or impediments" (p. 12).

Six Structure and Design standards are presented because "the design for any evaluation cannot be conceived in a vacuum. It is necessarily influenced by logistical, ethical, political, and fiscal concerns and therefore must take these as well as methodological requirements into account" (p. 13).

The section on Data Collection and Preparation includes twelve standards. These standards are based on the assumption a sound design and work plan have been developed. However, circumstances can change and these activities might need to be altered to reflect those changes.

Nine standards address Data Analysis and Interpretation. They are also based on the assumption a sound design and work plan have been developed, but analyses must be tempered to reflect the data actually collected.

Ten standards address Communication and Disclosure. Good communication is important in order to

> clarify the nature of the program, the expectations for the evaluation, and even the type of evaluation required . . . ; to anticipate restrictions on release of results and potential conflicts of interest . . . ; to establish accountability for the effort . . . ; to secure the cooperation of parties involved in the program and the evaluation . . . ; and to distinguish objective findings clearly from opinion and interpretation. (pp. 15-16)

Finally, six standards address the Use of Results. These standards are based on the assumption, "the use of evaluation results cannot be guaranteed, of course, but it will be more likely if careful attention is given to the information needs of potential users of the results throughout all phases of the evaluation" (p. 16).

Comparisons have also been made between the Evaluation Research Society and the Joint Committee Standards (e.g., Braskamp & Mayberry, 1982; Cordray, 1982; Stufflebeam, 1982). A common conclusion has been the similarities in the issues addressed and the expectations for quality, by and large, outweigh the differences. This suggests a high degree of agreement about the standards of quality exists in the field of evaluation.

Other Standards. Other evaluation standards have been written for more specialized purposes or audiences. For example, the U.S. General Accounting Office (1978) has a set of standards for assessing social program impact evaluations; and the U.S. Department of Education (1981) has published criteria to help select funding proposals submitted to the Office of Special Education that have sound evaluation designs.

Content Analysis Standards

No published standards for content analysis practice comparable to those for evaluation practice exist. However, two criteria for judging the quality of content analysis efforts--reliability and validity--are mentioned by a number of authors (e.g., Andrén, 1981;

Berelson, 1952; Budd, Thorp, & Donohew, 1967; Carney, 1972; Holsti, 1969; Janis, 1965; Kaplan & Goldsen, 1965; Krippendorff, 1980; Stone, Dunphy, Smith, & Ogilvie, 1966). Krippendorff (1980) highlights the importance of reliability and validity by defining content analysis as "a research technique for making replicable and valid inferences from data to their context" (p. 21). The Joint Committee (1981, pp. 116-123) and the ERS (1982, pp. 13-14) also have standards that specify evaluations should be concerned with both reliability and validity, particularly when delineating and obtaining information. This obviously applies to any evaluation that also uses content analysis methods.

Holsti (1969, pp. 135-149), Krippendorff (1980, pp. 129-168), and Stone et al. (1966, pp. 211-225) provide classifications of reliability and validity for content analysis. These classifications draw heavily on the classifications presented by a joint committee of the American Psychological Association, American Educational Research Association, and National Council on Measurement in Education (APA/AERA/NCME) (1974, pp. 25-55), although the content analysis classifications for reliability are less faithful to the APA/AERA/NCME classifications than those for validity.

Reliability. Table 4 is used to summarize the classifications of reliability used by the APA/AERA/NCME and the content analysis sources in terms of the type of agreement they best represent. A brief description of each type of reliability follows.

Test-retest agreement represents comparisons over time in general. Stability emphasizes an individual's self-agreement over time.

Table 4

Classifications of Reliability by Type of Agreement Represented

Type of agreement	APA/AERA/NCME	Holsti	Krippendorff	Stone et al.
Test-retest	Comparisons over time		Stability	Category stability
Inter-rater		Individual reliability	Reproducibility	Coder reliability
Test-standard			Accuracy	
Intra-unit			Unit reliability	
Rater-group			Individual reliability	
Intra-category	Internal consistency	Category reliability	Single category reliability	Category consistency
Intra-category group			Conditional reliability	
Inference-inference				Interpretive reliability
Form-form	Comparability of forms			

Note. Sources for the reliability classifications above include APA/AERA/NCME (1974, pp. 48-55), Holsti (1969, pp. 135-142), Krippendorff (1980, pp. 129-154), and Stone et al. (1966, pp. 211-217).

Category stability emphasizes a category's total score agreement over time.

Inter-rater agreement reflects the extent to which the pool of raters agrees on the collection of category ratings. This has been called individual reliability, reproducibility and coder reliability by the content analysis authors.

Test-standard agreement reflects the extent to which the pool of raters agrees with an external standard. Krippendorff calls this accuracy. To the extent the standard is considered to be "correct" or "true" he considers it to be a measure of validity (1980, p. 131).

Intra-unit agreement reflects the extent to which the pool of raters agrees on the coding of a single unit. This has been called unit reliability.

Rater-group agreement reflects the extent to which any single rater agrees with the remainder of the pool of raters. This has been called individual reliability by Krippendorff.

Intra-category agreement reflects the extent to which all the units in a single category represent the same concept. This has been called internal consistency, category reliability, single category reliability, and category consistency.

Intra-category group agreement is the same as above except more than one and fewer than all of the categories are simultaneously considered. This has been called conditional reliability.

Inference-inference agreement reflects the extent to which the pool of raters agrees on the conclusions that can be drawn from the analysis. This has been called interpretive reliability.

Form-form agreement reflects the extent to which administrations of parallel test forms produce comparable results. The APA/AERA/NCME calls this comparability of forms while no content analysis sources include this type of reliability.

Validity. Table 5 is used to summarize the classifications of validity used by the APA/AERA/NCME and the content analysis sources. Two of the sources simply adopt the APA/AERA/NCME classifications. Krippendorff relabels, subdivides, and groups the types of validity in terms of their orientation toward data, product, or process.

Content validity represents the extent to which a coding system and sample of units is representative of the universe of possible units. Krippendorff subdivides the representativeness of the coding system and sample as semantical validity and sampling validity.

Predictive validity represents the extent to which the results of a content analysis can be used to predict an event in the future. Concurrent validity represents the extent to which the results of a content analysis can be substituted for a different analytical procedure. Krippendorff calls both of these predictive validity. This is not consistent with the APA/AERA/NCME interpretation of predictive validity.

Construct validity represents the extent to which theoretical constructs have been substantiated. It requires multiple methods over multiple studies. Krippendorff includes two methods, convergent validation and discriminant validation, under the term correlational validity. Convergent validation represents the extent to which similar variables are highly and positively correlated. Discriminant

Table 5

Classifications of Validity

APA/AERA/NCME	Holsti	Krippendorff	Stone et al.
		Data oriented	
Content	Content	Semantical	Content
Content	Content	Sampling	Content
		Product oriented	
Construct	Construct	Correlational	Construct
Predictive	Predictive	Predictive	Predictive
Concurrent	Concurrent	Predictive	Concurrent
		Process oriented	
Construct	Construct	Construct	Construct

Note. Sources for the validity classifications above include APA/AERA/NCME (1974, pp. 25-48), Holsti (1969, pp. 142-149), Krippendorff (1980, pp. 155-168), and Stone et al. (1966, pp. 217-225).

validation represents the extent to which different variables are weakly correlated or negatively correlated. Krippendorff also calls construct validity the extent to which an analytical process parallels or mimics relations in the context where the data were created.

Summary

Standards of quality are important to the fields of evaluation and content analysis. However, evaluation standards are more formal than content analysis standards, even though these standards are best thought of as still emerging from multiple perspectives.

In addition, evaluation standards are a superset of content analysis standards--they both include expectations of reliability and validity while evaluation standards encompass a much wider range of expectations as well. Because of this, evaluation standards should always be applied whenever a study involves both characterizations and appraisals. If the study involves only descriptions and impartial conclusions, it does not constitute an evaluation. In this case, the standards of reliability and validity alone might suffice.

Working Definitions of Evaluation and Content Analysis

The link between evaluation and content analysis can now be established through working definitions that emphasize the similarities and differences between them in terms of information, actions, and standards of quality. Before the definitions themselves are presented, the key concepts on which they are based are reviewed.

First, evaluation and content analysis information focuses on different underlying contrasts. Evaluation information focuses on the contrast between knowledge and value statements. As a result, such information has been called characterizations and appraisals. Content analysis information focuses on the contrast between observations and logic. As a result, such information has been called descriptions and conclusions.

Second, evaluation and content analysis efforts involve the same basic actions. Four of these actions are related to processing some kind of information. They include delineating, obtaining, providing, and applying information. Two of these actions are related to the

total effort. They include managing and evaluating the evaluation or content analysis effort.

Third, high quality is important to both evaluation and content analysis, but the actual standards of quality are highly informal in content analysis and still emerging in evaluation. Because of this, normative definitions that simply draw attention to the issue of quality will be more durable than those that specify particular expectations of quality. Such definitions are sufficient here.

Based on the above considerations, working definitions of evaluation and content analysis with comparable grammatical structures follow. Good evaluation is the high quality process of delineating, obtaining, providing, and applying characterizations and appraisals about some object; and managing and evaluating the evaluation. Good content analysis is the high quality process of delineating, obtaining, providing, and applying descriptions and conclusions about some object; and managing and evaluating the content analysis.

An Overview of Evaluation

Several approaches to conducting evaluation efforts have been developed over the years. Many of these approaches make truly unique contributions to solving important problems, while others are little more than old goods in a new package. Classification systems intended to sort out unique approaches from repackaged goods are presented here to help identify some basic schools of thought for conducting an evaluation. After these approaches are identified, they are summarized in terms of a few important attributes.

Over the past twenty years or so, the number of alternative approaches to conducting evaluation efforts has skyrocketed. Factors such as the Elementary and Secondary Education Act (ESEA) of 1965 that required educators to evaluate their efforts and results, and the growing public concern for accountability of human service programs contributed to this growth (Borich & Jemelka, 1982, p. 4). Many sources that provide historical accounts of the proliferation of evaluation approaches are available. Works emphasizing educational evaluation include Baker (1980); Cronbach et al. (1980); Guba and Lincoln (1981); and Madaus, Stufflebeam, and Scriven (1983). Works on human service evaluation include Attkisson and Broskowski (1978), and Flaherty and Morell (1978). Borich and Jemelka (1982) provide a history of evaluation covering education and human services. Systems for classifying evaluation approaches are discussed next.

Classification of Approaches

A number of authors have provided classifications of various types of evaluation approaches (e.g., Borich & Jemelka, 1982, pp. 7-18; Guba & Lincoln, 1981, pp. 1-38; House, 1978; Patton, 1981, pp. 186-193; Popham, 1975, pp. 20-44; Stake, 1974; Stufflebeam & Webster, 1980; Worthen & Sanders, 1973, pp. 209-217). The number of general approaches presented in these sources covers a rather large range. For example, Guba and Lincoln (1981, p. 38) contend all evaluation approaches can ultimately be subsumed under one approach--responsive evaluation. At the other extreme, Patton (1981, pp. 186-193)--with tongue in cheek--presents a "beginning list" of 132 approaches. Most

other authors present around ten generalized evaluation approaches.

Two classifications of evaluation approaches (House, 1978; Stufflebeam & Webster, 1980) are of particular interest for this study because they include a manageable number of approaches, and they group these approaches in terms of underlying principles similar to those used in the preceding section of this chapter. The general structures of these classification systems are discussed first. The structures are then combined to present a more refined classification of fifteen evaluation approaches. A summary of these approaches is presented in the next section.

House (1978) considers all major evaluation approaches to be based on a common ideology, liberal democracy. Important principles of this ideology include freedom of choice, the uniqueness of the individual, and empirical inquiry. He also contends they are all based on subjectivist ethics, in which ethical conduct is based on the subjective or intuitive experience of an individual or group. One form of subjectivist ethics is utilitarian, in which "the good" is determined by what maximizes some single, explicit interpretation of happiness for society as a whole. Another form of subjectivist ethics is intuitionist/pluralist, in which no single interpretation of "the good" is assumed and these interpretations need not be explicitly stated nor justified.

These ethical positions have corresponding epistemologies--philosophies of obtaining knowledge. The objectivist epistemology is associated with the utilitarian ethic. In general, it is used to acquire knowledge capable of external verification (intersubjective

agreement) through publicly inspectable methods and data. The subjectivist epistemology is associated with the intuitionist/pluralist ethic. It is used to acquire new knowledge based on existing personal knowledge and experiences that are (explicit) or are not (tacit) available for public inspection.

House further divides each epistemological approach by two main political perspectives. Approaches can take an elite perspective, focusing on the interests of managers and professionals. They can also take a mass perspective, focusing on consumers and participatory approaches.

Stufflebeam and Webster (1980) place approaches into one of three groups according to their orientation toward the role of values, an ethical consideration. The political orientation promotes a positive or negative view of an object regardless of what its value might actually be. They call this pseudo-evaluation. The questions orientation includes approaches that might or might not provide answers specifically related to the value of an object. They call this quasi-evaluation. The values orientation includes approaches primarily intended to determine the value of some object. They call this true evaluation.

Table 6 is used to classify fifteen evaluation approaches in terms of epistemology, major perspective (from House), and orientation (from Stufflebeam & Webster). When considered simultaneously, these three dimensions produce twelve cells. Only seven of the cells contain approaches, although all four true evaluation cells contain at least one approach.

Table 6

Classification of Approaches for Conducting Evaluations Based on Epistemology, Major Perspective, and Orientation

Epistemology	Major perspective	Orientation		
		Political (Pseudo-evaluation)	Questions (Quasi-evaluation)	Values (True evaluation)
Objectivist (Utilitarian)	Elite (Managerial)	Politically controlled Public relations	Experimental research Management information systems Testing programs Objectives-based Content analysis	Decision-oriented Policy studies
	Mass (Consumers)		Accountability	Consumer-oriented
Subjectivist (Intuitionist/ Pluralist)	Elite (Professional)			Accreditation/ Certification Connoisseur
	Mass (Participatory)			Adversary Client-centered

Note. Epistemology and major perspective from House (1978). Orientation from Stufflebeam & Webster (1980).

Two pseudo-evaluation approaches, politically controlled and public relations studies, are represented. They are based on an objectivist epistemology from an elite perspective.

Six quasi-evaluation approaches use an objectivist epistemology. Five of them--experimental research, management information systems, testing programs, objectives-based studies, and content analysis--take an elite perspective. Accountability takes a mass perspective.

Seven true evaluation approaches are included. Two approaches, decision-oriented and policy studies, are based on an objectivist epistemology from an elite perspective. Consumer-oriented studies are based on an objectivist epistemology from a mass perspective. Two approaches--accreditation/certification and connoisseur studies--are based on a subjectivist epistemology from an elite perspective. Finally, adversary and client-centered studies are based on a subjectivist epistemology from a mass perspective. A summary of the fifteen approaches is presented next.

Summary of Approaches

The preceding section was used to distinguish between fifteen evaluation approaches in terms of their epistemology, major perspective, and orientation to values. This classification resulted in a twelve-celled matrix of which seven cells contain at least one entry. Five of the cells contain more than one entry. This section is intended to summarize each of the fifteen approaches in enough detail so that those placed in the same cell of Table 6 can be distinguished from each other.

Table 7 is used to summarize each approach in terms of four attributes--organizer, purpose, strengths, and weaknesses. The organizer represents the main considerations or cues practitioners use to organize a study. The purpose represents the desired outcome for a study at a very general level. Strengths and weaknesses represent other attributes that should be considered when deciding whether to use the approach for a particular study. Space considerations preclude most of the information in this table from being duplicated in the text. Instead, the following narrative highlights differences between approaches that are grouped into the same cell of Table 6. Sources for further reading on each approach are also presented here.

Pseudo-Evaluation. Only one of four pseudo-evaluation cells contains any entries. Politically controlled and public relations studies are based on an objectivist epistemology from an elite perspective. Although both of these approaches seek to misrepresent value interpretations about some object, they go about it a bit differently. Information obtained through politically controlled studies is released or withheld to meet the special interests of the holder. Public relations information is used to paint a positive image of an object regardless of the actual situation. Neither of these approaches is acceptable evaluation practice, although the seasoned reader can surely think of a few examples where they have been used.

Objectivist, Elite, Quasi-Evaluation. Five approaches are contained in this cell. As a group, they represent a highly respected collection of disciplined inquiry approaches. They are considered

Table 7

Summary of Approaches for Conducting Evaluations

	Attribute			
Approach	Organizer	Purpose	Key strengths	Key weaknesses
Politically controlled	Threats	Get, keep or increase influence, power, or money.	Secures evidence advantageous to the client in a conflict.	Violates the principle of full & frank disclosure.
Public relations	Propaganda needs	Create positive public image.	Secures evidence most likely to bolster public support.	Violates the principles of balanced reporting, justified conclusions, & objectivity.
Experimental research	Causal relationships	Determine causal relationships between variables.	Strongest paradigm for determining causal relationships.	Requires controlled setting, limits range of evidence, focuses primarily on results.
Management evidence systems	Scientific efficiency	Continuously supply evidence needed to fund, direct, & control programs.	Gives managers detailed evidence about complex programs.	Human service variables are rarely amenable to the narrow, quantitative definitions needed.

Table 7 (continued)

	Attribute			
Approach	Organizer	Purpose	Key strengths	Key weaknesses
Testing programs	Individual differences	Compare test scores of individuals & groups to selected norms.	Produces valid & reliable evidence in many performance areas. Very familiar to public.	Data usually only on testee performance, overemphasizes test-taking skills, can be poor sample of what is taught or expected.
Objectives-based	Objectives	Relate outcomes to objectives.	Common sense appeal, widely uses, used behavioral objectives & testing technologies.	Leads to terminal evidence often too narrow to provide basis for judging the value of a program.
Content analysis	Content of a communication	Describe & draw conclusions about a communication.	Allows for unobtrusive analysis of large volumes of unstructured, symbolic materials.	Samples may be unrepresentative yet overwhelming in volume. Analysis design often overly simplistic for question.
Accountability	Performance expectations	Provide constituents with an accurate accounting of results.	Popular with constituents. Aimed at improving quality of products and services.	Creates unrest between practitioners & consumers. Politics often forces premature studies.

Table 7 (continued)

	Attribute			
Approach	Organizer	Purpose	Key strengths	Key weaknesses
Decision-oriented	Decisions	Provide a knowledge & value base for making & defending decisions.	Encourages use of evaluation to plan & implement needed programs. Helps justify decisions about plans & actions.	Necessary collaboration between evaluator & decision-maker provides opportunity to bias results.
Policy studies	Broad issues	Identify and assess potential costs & benefits of competing policies.	Provides general direction for broadly focused actions.	Often corrupted or subverted by politically motivated actions of participants.
Consumer-oriented	Generalized needs & values, effects	Judge the relative merits of alternative goods & services.	Independent appraisal to protect practition- & consumers from shoddy products & services. High public credibility.	Might not help practitioners do a better job. Requires credible & competent evaluator.
Accreditation/certification	Standards & guidelines	Determine if institutions, programs, & personnel should be approved to perform specified functions.	Helps public make informed decisions about quality of organizations & qualifications of personnel.	Standards & guidelines typically emphasize intrinsic criteria to the exclusion of outcome measures.

Table 7 (continued)

	Attribute			
Approach	Organizer	Purpose	Key strengths	Key weaknesses
Connoisseur	Critical guideposts	Critically describe, appraise, & illuminate an object.	Exploits highly developed expertise on subject of interest. Can inspire others to more insightful efforts.	Dependent on small number of experts, making evaluation susceptible to subjectivity, bias, and corruption.
Adversary	"Hot" issues	Present the pros & cons of an issue.	Ensures balanced presentations of represented perspectives.	Can discourage cooperation, heighten animosities.
Client-centered	Specific concerns & issues	Foster understanding of activities & how they are valued in a given setting & from a variety of perspectives.	Practitioners are helped to conduct their own evaluation.	Low external credibility, susceptible to bias in favor of participants.

Note. Adapted and condensed primarily from House (1978) and Stufflebeam & Webster (1980).

quasi-evaluation approaches because particular studies can legitimately focus only on questions of knowledge without addressing any questions of value. Such studies are, by definition, not evaluations. Using the terminology developed in the first section of this chapter, these approaches can produce characterizations without producing appraisals, although specific studies can produce both. Each of these approaches serves its intended purpose well. They are discussed roughly in order of the extent to which they approach the objectivist ideal.

Experimental research is the best approach for determining causal relationships between variables. The potential problem with using this as an evaluation approach is that its highly controlled and stylized methodology may not be sufficiently responsive to the dynamically changing needs of most human service programs. Important contributors to the experimental research approach include Lindquist (1953), and Cook and Campbell (1979).

Management information systems (MIS's) can give detailed information about the dynamic operations of complex programs. However, this information is restricted to readily quantifiable data usually available at regular intervals. Contributors to the MIS approach include Cook (1966), Kauffman (1969), and Rivlin (1971).

Testing programs are familiar to just about anyone who has attended school, served in the military, or worked for a large company. These programs are good at comparing individuals or groups to selected norms in a number of subject areas or to a set of standards of performance. However, they only focus on testee performance and

they might not adequately sample what is taught or expected. Contributors to the testing program approach include Lindquist (1951), Ebel (1965), Hambleton and Swaminathan (1985), and Thorndike (1971).

Objectives-based approaches relate outcomes to prespecified objectives, allowing judgments to be made about their level of attainment. Unfortunately, the objectives are often not proven to be important or they focus on outcomes too narrow to provide the basis for determining the value of an object. Contributors to this approach include Tyler (1949); Bloom, Englehart, Furst, Hill, and Krathwohl (1956); Hammond (1973); Kiresuk and Lund (1978); Krathwohl, Bloom, and Masia (1964); Metfessel and Michael (1967); Popham (1969); and Provus (1971).

Content analysis was not included in the original classifications of evaluation approaches used for this section (House, 1978; Stufflebeam & Webster, 1980). It was added to the list to place it in the context of commonly accepted evaluation approaches. It is a quasi-evaluation approach because content analysis judgments need not be based on value statements. Instead, they can be based on knowledge. Such content analyses are not evaluations. On the other hand, when content analysis judgments are based on values, such studies are evaluations. This approach is discussed in detail in a later section. Key contributors to content analysis include Berelson (1952); Krippendorff (1980); Lasswell, Leites, and Associates (1965); Holsti (1969); and Stone et al. (1966).

Objectivist, Mass, Quasi-Evaluation. Accountability is the only approach assigned to this cell. It is popular with constituents

because it is intended to provide an accurate accounting of results that can improve the quality of products and services. However, this approach can quickly turn practitioners and consumers into adversaries when implemented in a heavy-handed fashion. The leading contributor to the accountability approach is Lessinger (1970).

Objectivist, Elite, True Evaluation. The evaluation approaches in this cell are of two basic types. They include decision-oriented studies and policy studies.

Decision-oriented studies are designed to provide a knowledge base for making and defending decisions. This approach usually requires the close collaboration between an evaluator and decision-maker, allowing it to be susceptible to corruption and bias. Contributors to this approach include Cronbach (1963), Stufflebeam et al. (1971), and Alkin (1969).

Policy studies provide general guidance and direction on broad issues by identifying and assessing potential costs and benefits of competing policies. The drawback is these studies can be corrupted or subverted by the politically motivated actions of the participants. Contributors to the policy study approach include Coleman et al. (1966), Jenks et al. (1972), and Clark (1965).

Objectivist, Mass, True Evaluation. One approach has been placed in this cell, consumer-oriented studies. This approach is used to judge the relative merits of goods and services based on generalized needs and values, along with a comprehensive range of effects. However, this approach does not necessarily help practitioners improve their work, and it requires a very good and credible

evaluator to do it well. The single most important contributor to the description and practice of this evaluation approach is Scriven (1967, 1974b).

Subjectivist, Elite, True Evaluation. This cell contains three approaches, accreditation/certification, policy studies, and connoisseur studies. They represent one of two groups of approaches that use a subjectivist epistemology.

Accreditation/certification programs are based on self-study and peer review of organizations, programs, and personnel. They draw on the insights, experience, and expertise of qualified individuals who use established guidelines to determine if the applicant should be approved to perform specified functions. However, attributes of applicants and the processes they perform are often overemphasized in relation to measures of outcomes or effects. Examples of accreditation boards include the Commission on Accreditation of Rehabilitation Facilities, the Joint Commission on Accreditation of Hospitals, and the North Central Association of Secondary Schools and Colleges.

Connoisseur studies use the highly refined skills of individuals intimately familiar with the subject of the evaluation to critically characterize and appraise it. This approach can help others see programs in a new light, but it is difficult to find a qualified and unbiased connoisseur. Contributors to this approach include Eisner (1975), Guba (1978), and Sanders and Hershiser (1976).

Subjectivist, Mass, True Evaluation. This final cell contains two approaches, adversary and client-centered studies. They use a subjectivist epistemology from multiple audience perspectives.

The adversary approach focuses on drawing out the pros and cons of controversial issues through quasi-legal proceedings. This helps ensure a balanced presentation of different perspectives on the issues, but it is also likely to discourage later cooperation and heighten animosities between contesting parties if "winners" and "losers" emerge. Contributors to this approach include Owens (1971) and Wolf (1973).

The last approach covered includes client-centered studies. They address specific concerns and issues of practitioners and other clients of the study in a particular setting. These studies help people understand the activities and values involved from a variety of perspectives. However, this responsive approach can lead to low external credibility and a favorable bias toward those who participated in the study. Contributors to this approach include Stake (1967), Guba (1978), Guba and Lincoln (1981), and Rippey (1973).

A General Model for Conducting an Evaluation Effort

The first area of discussion in this chapter focused on the conceptual relationships between evaluation and content analysis. The second area of discussion focused on alternative approaches to conducting evaluations and identified content analysis as a quasi-evaluation approach. This third area of discussion focuses on a general model for conducting an evaluation effort that is consistent with the previous discussions; accommodates content analysis experts' views on the uses, tasks, and computer implementations of content analysis; and provides the framework for an experimental study on how

microcomputers can be used to improve content analyses of responses to open-ended survey questions used in evaluation efforts.

The general model for conducting an evaluation effort reflects the key relationships between the information, action, and standards of quality components discussed earlier. It can also be thought of as a graphic version of the working definition of evaluation. The model is presented in Figure 2.

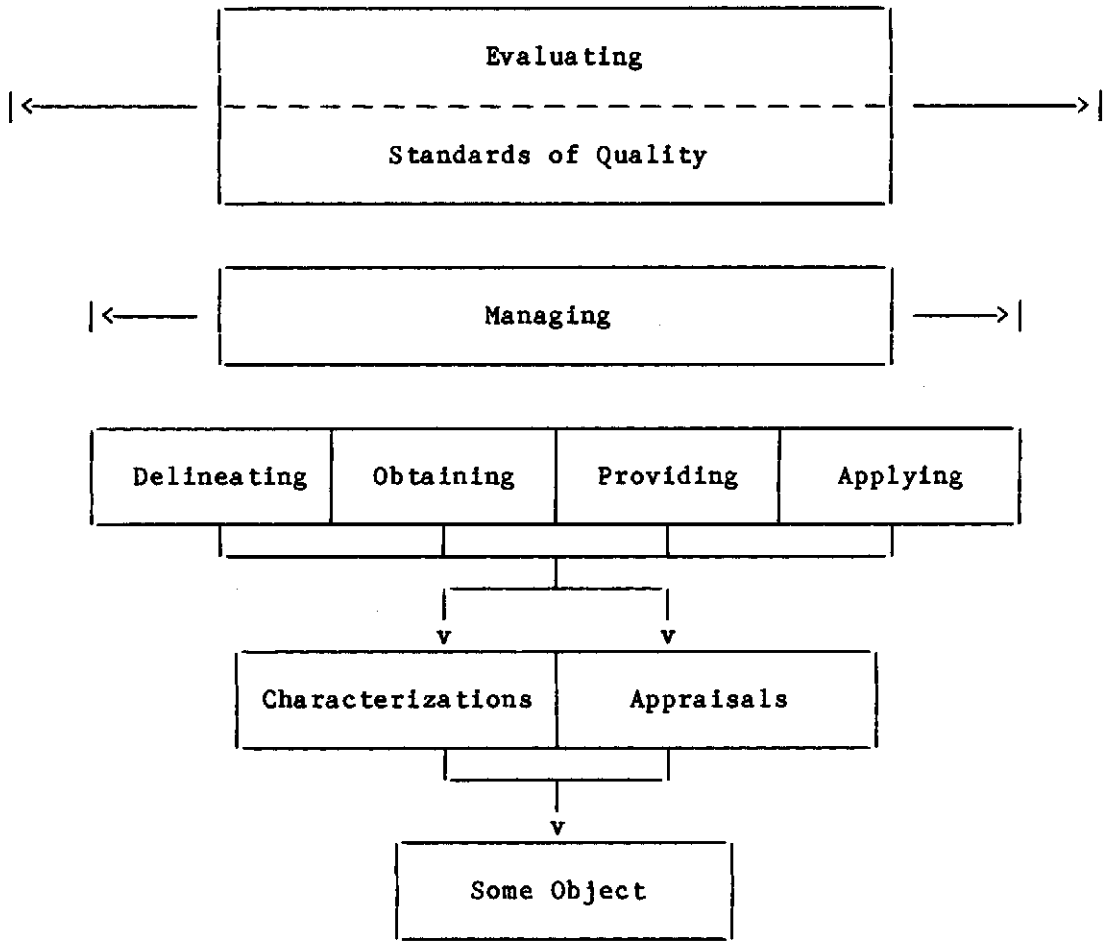


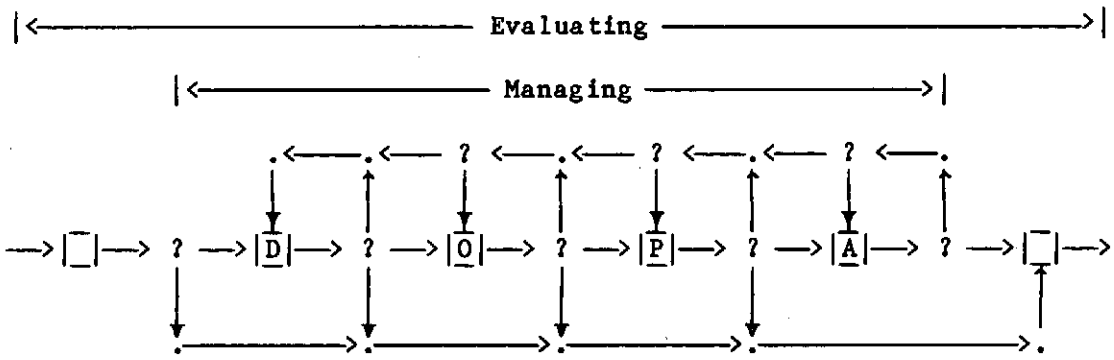
Figure 2. The General Model for Conducting an Evaluation Effort

In this context, evaluation is best thought of as a process that includes six actions. Four of them--delineating, obtaining, providing, and applying--focus on information. This information includes characterizations and appraisals about some object.

The other two basic actions--managing and evaluating--focus on the overall effort. Managing the effort begins at some time during the early delineating activities and it ends at some phase of applying the information. Exactly when management of the effort begins and ends is dependent on the particular evaluation approach used. Evaluating the effort, or meta-evaluation, can be used to scrutinize events that occurred well before and after the official time period of the study, although most of the focus is usually placed on the official information processing actions and their consequences. It should also be noted the working definition stipulates all the actions and information should be of high quality, including managing and evaluating the effort.

While Figure 2 is used to show the basic relationships between the information, actions, and standards of quality for an evaluation effort, Figure 3 better illustrates its dynamic nature. This figure is used to focus on the decision network for delineating, obtaining, providing, and applying evaluative information--characterizations and appraisals. The network itself is represented in the lower portion of the figure, while the upper portion is used to remind the reader these actions still need to be managed and evaluated.

From a logical perspective, any path that follows the flow of the arrows is possible, although certain patterns are more likely



Legend. |D| = Delineating information, |O| = Obtaining information, |P| = Providing information, |A| = Applying information, |_| = Unspecified action, ? = Which action next?

Figure 3. Decision Network Using the General Model for Conducting an Evaluation Effort

than others. For example, if one were to ask if an evaluation should be conducted at all but immediately answered no, the lower path that bypasses the heart of the process is followed.

This network can also accommodate two ideal types of evaluation information processing patterns that are conceptually incompatible but are probably never found in their "pure" forms. These patterns of information processing are often called "preordinate" and "responsive" in the evaluation literature (e.g., Guba & Lincoln, 1981; Stake, 1975). The preordinate pattern is exemplified by the experimental research approach presented earlier. In this pattern, all delineating activities are completed before any of the obtaining activities begin. In a like manner, all obtaining activities are completed before providing information begins, which is completed

before applying information begins. In the responsive pattern, exemplified by the client-centered approach discussed earlier, several iterations of delineating, obtaining, providing, and applying information about each aspect and subplot of an evaluation are undertaken before the effort as a whole is completed. In reality, however, even preordinate efforts often need to follow side issues or return to a previous stage of a study to modify work already completed; and responsive efforts often complete substantial portions of a particular type of action before moving on to the next stage of the study.

In addition, this network is hierarchically recursive in nature. In other words, in order to complete a major information processing action, supporting actions often need to be completed first. For example, before providing an appraisal of a school district's accountability system to the school board, it is first necessary to characterize how various interest groups view the system.

In summary, the general model for conducting an evaluation effort can be used to show the logical relationships between its information, action, and standards of quality components. It can also accommodate many different patterns of processing characterizations and appraisals of some object.

Because content analysis is a quasi-evaluation approach, the general model also applies to such efforts. The remainder of this chapter focuses on that approach in terms of its general uses, tasks, and the performance of certain of these tasks with the help of microcomputers.

An Overview of Content Analysis

Around the turn of the century, the use of content analysis as a serious approach to studying the world in general and the nature of communication in particular began to increase. Because most early studies were concerned with describing daily newspapers in terms of the amount of coverage they gave to different subject areas (e.g., Speed, 1893), the approach was then called "quantitative newspaper analysis" (Krippendorff, 1980, p. 14). However, the popularity of the approach soon began to grow into different fields (e.g., military intelligence, psychotherapy, history, anthropology, education, literature, & linguistics) and different media (e.g., books, documents, correspondence, movies, radio, TV, & photography). Historical accounts of this growth in the field are available from a number of sources (e.g., Barcus, 1959; Berelson, 1952, pp. 21-26; Carney, 1972, pp. 26-36; Holsti, 1969, pp. 20-23; Krippendorff, 1980, pp. 13-20; Stone et al., 1966, pp. 21-44). This discussion is organized into three basic topics: (1) a summary of the uses of content analysis, (2) a summary of the tasks of content analysis, and (3) a summary of the ways computers in general and microcomputers in particular can be used to help perform some of these tasks.

Classification of Uses

Several authors have provided classifications of content analysis and its uses. One of the earliest classification systems was developed by Janis (1943). This classification system also appears

in a book devoted to using content analysis methods to help study the field of politics (Janis, 1965). The three main types of content analysis are described as:

1. Pragmatical Content Analysis--procedures which classify signs according to their probable causes or effects (e.g., counting the number of times that something is said which is likely to have the effect of producing favorable attitudes toward Germany in a given audience).
2. Semantical Content Analysis--procedures which classify signs according to their meaning (e.g., counting the number of times Germany is referred to, irrespective of the particular words that may be used to make the reference).
   (a) designations analysis--provides the frequency with which certain objects (persons, things, groups or concepts) are referred to, i.e., roughly speaking, subject-matter analysis (e.g., references to German foreign policy).
   (b) attribution analysis--provides the frequency with which certain characterizations are referred to (e.g., references to dishonesty).
   (c) assertions analysis--provides the frequency with which certain objects are characterized in a particular way, i.e., roughly speaking, thematic analysis (e.g., references to German foreign policy as dishonest).
3. Sign-vehicle analysis--procedures which classify content according to the psychophysical properties of the signs (e.g., counting the number of of times the word "Germany" appears). (p. 57)

Berelson (1952) places seventeen uses of content analysis into four basic groups, one of which has two subgroups. They are: "[1] characteristics of content . . . , [a] substance . . . , [b] form . . . ; [2] producers of content . . . ; [3] audiences of content . . . ; and [4] effects of content" (pp. 27-29).

Holsti (1969) draws from the above sources to group fifteen uses into three basic categories. His three categories are to: "[1] describe characteristics of communication . . . , [2] make inferences

as to the antecedents of communication . . . , and [3] make inferences as to the effects of communication" (p. 26). He also places these uses into the context of the communication paradigm and states:

> Content analysis is always performed on the message, be it a novel, diplomatic note, editorial, diary, or speech. The results of content analysis may, however, be used to make inferences about all other elements of the communication process. To the classical formulation of these questions--"who says what, to whom, how, and with what effect?" (Lasswell, Lerner, and Pool, 1952, p. 12)--we shall add one more: "why?" (p. 24)

Although the details of Krippendorff's (1980) discussion on the types of content analysis are essentially the same as those of the above authors, he groups them differently into six main categories. His categories include "[1] systems . . . , [2] standards . . . , [3] indices and symptoms . . . , [4] linguistic representations . . . , [5] communications . . . , and [6] institutional processes" (p. 34).

Because Holsti's classification system accommodates almost all of the specific uses presented by the other authors, its underlying structure--the communication paradigm--is familiar to a general audience, and it is compatible with the working definition of content analysis; it is used as the main organizer for the discussion on the uses of content analysis. The contributions of the other authors that expand this basic framework are added where appropriate.

Summary of Uses

Table 8 is used to summarize fifteen uses of content analysis in terms of their general purpose, element of the communication paradigm

Table 8

Uses of Content Analysis by Purpose, Communication Element, and Question

Purpose	Element	Question	Use
Make inferences about the antecedents of communications	Source	Who?	Answer questions of disputed authorship
	Encoding process	Why?	Secure political & military intelligence
			Analyze traits of individuals
			Infer cultural aspects & change
			Provide legal & evaluative evidence
Describe & make inferences about the characteristics of communications	Channel	How?	Analyze techniques of persuasion
			Analyze style
	Message	What?	Describe trends in communication content
			Relate known characteristics of sources to messages they produce
			Compare communication content to standards
	Recipient	To whom?	Relate known characteristics of audiences to messages produced for them
			Describe patterns of communication
Make inferences about the consequences of communications	Decoding process	With what effect?	Measure readability
			Analyze the flow of information
			Assess responses to communications

Note. Purpose, communication element, & question from Holsti (1969). Uses primarily from Berelson (1952) as adapted by Holsti (1969).

to which they apply, and the general question they are intended to answer. In Figure 4, this same information is graphically presented in relation to the working definition of content analysis. Thus, using action forms of the definition's information components, content analysis is a process for describing messages and making inferences about any element of the communication process. Managing and evaluating the effort are also necessary if high quality results are desired. The uses of content analysis that apply to the antecedents, characteristics, and consequences of communications are discussed next. However, due to space considerations, references to specific content analysis studies and methods are kept to a minimum. Readers interested in learning more about particular uses in specific contexts are referred to Berelson (1952) and Holsti (1969).

Antecedents of Communications

Making inferences about the antecedents of communications is a problem in pragmatics--identifying the relationships of signs (messages) to those who produced them. Because of differences in the ways people express their feelings, intentions, and other attributes,

> inferences about the antecedent causes of messages drawn solely from content data cannot be self-validating. Thus, however precise our measures of communication content, it is hazardous indeed to assume, without corroborating evidence from independent, noncontent data, that inferences about the author may be drawn directly from content data. (Holsti, 1969, p. 32)

Because of this problem, two basic types of comparisons--direct and indirect--are used to draw inferences about the antecedents of

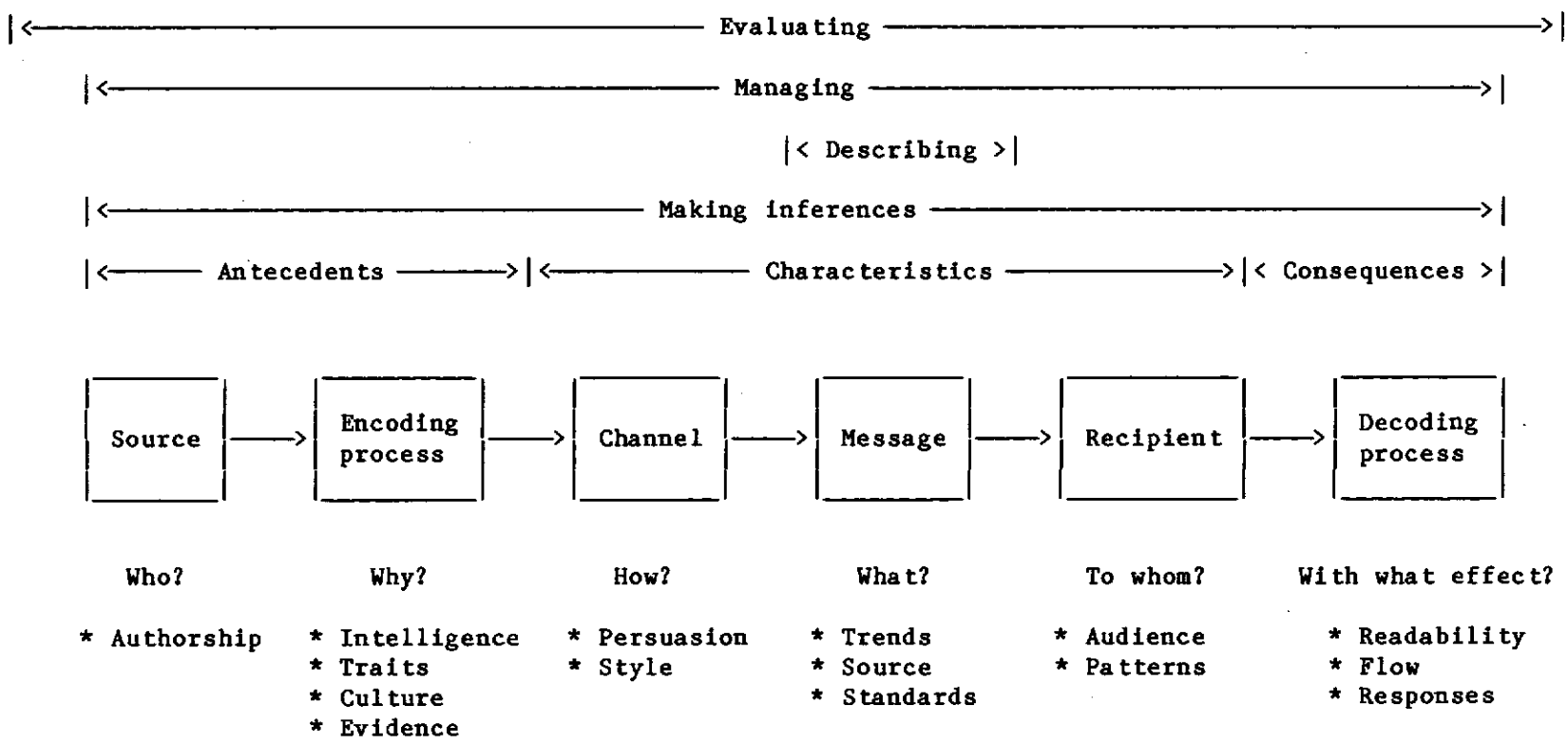


Figure 4. Uses of Content Analysis in Terms of the Working Definition and the Communication Paradigm

messages. With direct comparisons, a content variable from a given source (e.g., aggressive threats of a specific terrorist group) is compared to a behavioral variable of the same type for the same source (e.g., actual terrorist attacks by that group).

Indirect comparisons, on the other hand, relate a content variable from one source to the same content variable from a different source. The different source can be a completely different individual or the same individual at a different point in time. In addition, the content variable from the other source must have been directly compared to a related behavioral variable. For example, comparing the coverage of one textbook to that of another considered to be culturally biased can be used to make inferences about the biases of the first textbook. An example of an indirect comparison for a single individual is when a mother sends her four-year-old to bed at 7:30 P.M. after hearing a series of complaints--knowing full well things will only get worse if the child is allowed to stay up any later.

Any inferences about the antecedents of a communication must also be based on either a representational or instrumental model of communication. In the representational model,

> the important point about the communication is what is revealed by the content of the [words] present in it; that is, something in the words of the message may have indicatorial validity regardless of the circumstances, and it is at the message that the analyst looks. (Pool, 1959, p. 3)

The instrumental model "signifies that the important point is not what the message says on the face of it but what it conveys, given

its context and circumstances" (p. 3). For a further discussion of the representational and instrumental models, see Pool (1959, chaps. 1-3, 7) and Mitchell (1967).

Two questions--"who?" and "why?"--apply to making inferences about the antecedents of communications. "Who?" refers to the source of a message. Answering questions of disputed authorship is the only use of content analysis included here. "Why?" refers to the encoding process for a message. "What are the meanings, associations, values, motives, or intentions of the communicator that can be inferred from his [or her] messages?" (Holsti, 1969, p. 32). The uses of content analysis in this group are to: (a) secure potential military intelligence, (b) analyze psychological traits of individuals, (c) infer cultural aspects and change, and (d) provide legal or evaluative evidence.

Two uses, (b) and (d), are probably more interesting to evaluators than the others. For example, surveys can be used to collect evidence about attitudes different interest groups have concerning the object of an evaluation. When they use open-ended questions--questions for which people are asked to respond in their own words--content analysis methods can be used to help summarize the narrative responses (Berelson, 1952, pp. 53-56; Caulley, 1983, p. 22). Methods for analyzing such responses and some suggestions for when open-ended questions should be used at all are discussed next.

Evaluative assertion analysis (Osgood, 1959; Osgood, Saporta, & Nunnally, 1956) can be used to determine the attitudes of different groups about a particular evaluation object, such as a public

school's accountability system. "The purpose of evaluative assertion analysis is to extract from messages the evaluations being made [by the source] of significant concepts" (Osgood, 1959, p. 42). The crucial first phase of evaluative assertion analysis involves transforming the original message into a series of evaluative assertions ("neutral" assertions are not included in this particular type of analysis) with one of two possible grammatical structures: (1) attitude object / verbal connector / common meaning term, or (2) attitude $object_1$ / verbal connector / attitude $object_2$. For example, the sentence, "All teachers despise the dogmatic accountability system," is translated to read: (1) All teachers / despise / the accountability system, and (2) The accountability system / is / dogmatic. Quantitative techniques are then performed on these transformed assertions. The final result is a numerical indicator of the direction and intensity of all evaluative assertions about each attitude object. However, the transformed assertions themselves constitute a type of "qualitative" information that can also be very useful. Unfortunately, because of the labor-intensive nature of the method, Osgood (1959) concludes, "This method is more likely to find use as a research tool than in practically oriented areas" (p. 53). Evaluation qualifies as one of these practically oriented areas.

Another method developed by Osgood for summarizing the attitudes of reference groups is the semantic differential technique (Osgood, 1962; Osgood Suci, & Tannenbaum, 1957; Snider & Osgood, 1969). This method is <u>not</u> used to analyze narrative data. Instead, concepts or attitude objects are related in terms of a number of bipolar scales

like "aloof-responsive," or "fair-biased." These attitude objects are then represented in three empirically derived dimensions of meaning--evaluation (e.g., good-bad or positive-negative), potency (e.g., strong-weak or hard-soft), and activity (e.g., active-passive or slow-fast)--(Osgood et al., 1957, pp. 62-63). However, comparing narrative data to these dimensions, particularly evaluation, is possible. For example, responses to open-ended questions can first be coded by the main attitude object they contain and then each response can be scored on a rating scale (e.g., Stanley & Hopkins, 1972, p. 290) that represents the evaluation dimension of the semantic differential (e.g., positive-negative). Depending on the precision desired, the scale can have anywhere from two to seven points. Using a three-point scale (negative-neutral-positive) and the sentence, "All teachers despise the dogmatic accountability system;" the main attitude object is, "the accountability system," and the rating on the evaluation scale is, "negative."

Survey researchers usually distinguish between two basic types of questions--open-ended and forced-choice. Open-ended questions require people to respond in their own words while forced-choice questions require people to choose from a set of predefined alternatives. Obviously, it is the open-ended type of question whose responses must be subjected to content analysis.

Payne (1951) identifies several uses of open-ended questions. Common uses are to: (a) introduce a topic to the respondent, (b) provide background information for interpreting responses to other questions, (c) obtain elaborations of previous responses, (d) elicit

reasons for previous responses, (e) elicit presumed arguments for different sides of an issue, (f) explore knowledge and memory, (g) identify sources of information, (h) obtain factual information, (i) provide preliminary information for drafting forced-choice questions, and (j) provide a source of quotations for final reports (pp. 34-50).

Open-ended and forced-choice questions both have characteristic strengths and weaknesses (e.g., Demaline & Quinn, 1979, pp. 30-31; Downs, Smeyak, & Martin, 1980, pp. 44-48; Payne, 1951, pp. 49-54; Warwick & Linnger, 1975, pp. 132-140). The summary by Demaline and Quinn (1979) is both concise and representative of the other authors. They present the following advantages and disadvantages of open-ended and forced-choice questions:

Forced-Choice Questions

Advantages
1. It is easier for respondent to answer.
2. Focuses respondent's answer on issues and data of importance to you. Respondents categorize themselves instead of you categorizing them.
3. More questions can be asked because time is saved by the respondent simply checking.
4. Precoded answers are easily analyzed.

Disadvantages
1. It requires advance information about possible response categories that may be given.
2. It may bias responses by suggesting answers.
3. It does not allow for diversity and richness in individual expression.

Open-Ended Questions

Advantages
1. It can easily be formulated without knowing the full range of answers that may be given.
2. It can accommodate questions for which a wide range of different answers will be given.
3. It does not condition or bias the answer as much as the forced-choice question.

Disadvantages

1. It requires the respondent to write a lot. Communication skills may influence the answer, in addition [to] the other respondent characteristics.
2. The respondent may address different facets of the question in which you may not be interested or may not give complete information in answering the question.
3. Fewer questions can be asked in a questionnaire because answers may be lengthy and time consuming to give.
4. Responses are difficult to analyze. The investigator must devise a coding scheme and then categorize responses based on this scheme. The diversity and richness of responses are usually reduced by this process, and it is time consuming. (pp. 30-31)

Because the advantages and disadvantages of open-ended and forced-choice questions tend to complement each other, suggestions for when either type of question should be used can be given. For example, use forced-choice questions if any of these four conditions apply: (1) it is important the effort or verbal skills of the respondents be kept relatively low, or (2) a clear understanding exists of what the likely or important responses will be, or (3) a large number of questions need to be asked in relation to the time available, or (4) it is important the responses be easy to code for analysis. Use open-ended questions and content analysis if either of these two conditions apply: (1) the full range of likely and valid responses is not known or a wide range of responses is expected, or (2) concern for biasing respondents exists if a set of possible answers is given, and this third condition applies: (3) the skills and time needed for coding potentially complex responses are available. Thus, open-ended questions can be extremely useful, but the effort and content analysis skills needed to properly analyze the responses to them require their judicious use. Payne (1951)

concludes, "Its virtues and its faults all stem from this open feature. Its results are as full of variety as a country store, and just as hard to divide into departments" (p. 54).

Characteristics of Communications

Describing and making inferences about the characteristics of communications are problems in semantics and syntactics, what Janis (1965) calls semantical content analysis and sign-vehicle analysis (p. 57). Exactly when "describing" becomes "making inferences" is highly dependent on the analyst's position on the objectivist vs. subjectivist issue as discussed under the classification of alternative evaluation approaches. The representational vs. instrumental communication model controversy is a specific example of this general issue. Analysts using the representational model are more likely to "describe" the characteristics of communications, while analysts using the instrumental model are more likely to "make inferences" about the characteristics of communication.

Three basic questions--"how?", "what?", and "to whom?"--apply to describing and making inferences about the characteristics of communications. "How?" refers to the channel of a message. The uses of content analysis in this group are to: (a) analyze techniques of persuasion, and (b) analyze style. "What?" refers to the message itself. The uses grouped under this question are to: (a) describe trends in communication content, (b) relate known characteristics of sources to messages they produce, and (c) compare communication content to standards. The use of content analysis to compare

messages to standards is discussed next. "To whom?" refers to the recipient of the message. The content analysis uses here are to: (a) relate known characteristics of audiences to messages produced for them, and (b) describe patterns of communication.

Berelson (1952), Holsti (1969), and Krippendorff (1980) all partition the uses of content analysis related to standards differently. However, it should be noted they all use the term to mean what is compared to a message, not what is used to judge the quality of the content analysis descriptions, inferences, or processes used to develop them.

Berelson (1952) uses two main categories. The first one, to audit communications against objectives (pp. 43-45), checks a message "against the communicator's own professed objectives" (p. 44). The second category, to construct and apply communication standards (pp. 46-52), compares messages "with the standards of the analyst" (p. 44). He also lists three methods of evaluating communications:

> (1) evaluation of performance against such a priori standards as "balance" or "social purpose"; (2) evaluation of performance by comparing one body of content with another (the internal criteria); and (3) evaluation of performance by comparing content with a non-content source (the external criteria). (p. 46)

Holsti (1969) drops the distinction between the source of the standards but maintains the distinction between the three types of standards--a priori, content, and non-content (pp. 53-59). His examples of a priori standards used in actual studies include social norms, responsibility in mass communication, and bias (pp. 54-56).

Content standards are based on "general norms for classes of communicators" (p. 56). Non-content standards include such things as verified news events to assess newspaper story coverage (p. 54); census data to assess the distribution of ethnic group members in popular fiction (p. 58); and expert opinion as the basis to judge depictions of mental health issues in the general press, or adherence to the "American creed" in ethnic newspapers (p. 58).

Krippendorff (1980) identifies three types of content analysis studies that use standards--evaluations, identifications, and audits.

> While evaluations assess the degree [emphasis added] to which something conforms or deviates from a standard, identifications have a more either/or quality. . . . Audits, too, involve judgments on data relative to a standard with the additional provision that the standard is prescribed or legitimated by an institution. (p. 39)

Krippendorff's view of identifications most closely resembles Janis' (1965) categories of semantical and sign-vehicle content analysis (p. 57). Identifications also reflect the importance of the role of concept learning in content analysis. "Learning a concept means learning to classify stimulus situations in terms of abstracted properties like color, shape, position, number and others" (Gagné, 1970, p. 51). Individuals who understand a concept can do at least four things: (1) identify the concept when examples of it are presented, (2) state properties of the concept, (3) distinguish examples from nonexamples, and (4) better solve problems that include the concept (Davis, Alexander, & Yelon, 1974, p. 221). However, "the best demonstration of knowledge of the concept is the . . . ability

to differentiate examples from nonexamples" (p. 222). Put another way, "Knowing a concept involves being able to classify objects or events" (p. 221), a crucial task in any content analysis effort. It requires intellectual abilities and skills classified under application in the Taxonomy of Educational Objectives (Bloom et al., 1956, pp. 120-143). The reader is referred to Klausmeier & Harris (1966) for an anthology of papers on the topic of concept learning, and Winston (1984, chaps. 11 & 12) for a summary of computer-based, artificial intelligence implementations of concept learning.

Krippendorff's view of audits most closely resembles the quasi-evaluation approach called accountability in this chapter. His view of evaluation as a whole is characteristic of most content analysts in that it fails to distinguish between inferences based on knowledge and inferences based on value statements. This lack of distinction creates problems in justifying the standards selected. All three authors (Berelson 1952, p. 46; Holsti, 1969, p. 56; Krippendorff, 1980, p. 39) agree most content analysis studies that compare messages to a set of evaluative standards provide little or no justification for the selection of the standards used. This problem can be addressed by thinking of content analysis as a quasi-evaluation approach and using evaluation standards of quality (e.g., ERS Standards Committee, 1982; Joint Committee, 1981) when content analysis inferences produce determinations of merit or determinations of worth, rather than "simply" impartial conclusions.

Thus, content analysts have identified how messages can be compared to standards along three different dimensions. First, the

creators of the standards can be identified. The three main creators of standards are the producer of the message, the content analyst, and reference groups selected by the analyst. Second, the types of justifications for standards can be identified. Standards have been justified before the fact, because they reflect the normative characteristics of the applicable type of message, or because they are empirically derived from related, non-content sources. Third, the type of information they produce can be identified. Using the language of this study, comparisons of messages to standards grounded in concept learning produce characterizations. Comparisons of messages to standards grounded in value statements produce appraisals.

Consequences of Communications

Making inferences about the consequences of communications is a problem in pragmatics--identifying the relationships of signs (messages) to the effects they produce on the receivers of the messages. Using experimental design terminology, the message is the independent variable and the effect of the message is the dependent variable. In this case, the independent variable is measured through content analysis. The dependent variable can be measured in one of two basic ways: (1) content analysis of audience messages produced in response to the original message, or (2) any other behavioral measures that do not require content analysis.

Of course, factors besides the message of interest will also have an impact on its audience. For example, preexisting attitudes of the audience, the perceived credibility of the communicator, and

personality characteristics of the audience can all temper the effects of the message (Holsti, 1969, p. 88). Content analysis studies with adequate experimental or quasi-experimental designs (e.g., Campbell & Stanley, 1963; Cook & Campbell, 1979) can at least partially control for some of these factors. Holsti (1969) also notes, "Owing to the possible effects of factors other than message content, including audience predispositions and decoding habits, the effects of communication cannot be directly inferred from the attributes of content (what) or style (how) without independent validation" (p. 88).

One question--"with what effect?"--applies to making inferences about the consequences of communications. This question refers to the communication element of decoding the message. The uses of content analysis in this group are to: (a) measure readability, (b) analyze the flow of information from one source to another, and (c) assess responses to communications. The reader is referred to Berelson (1952, pp. 98-108) and Holsti (1969, pp. 87-93) for discussions of specific studies designed to make inferences about the consequences of communications.

Tasks of Content Analysis

Good content analysis has been defined as the high quality process of delineating, obtaining, providing, and applying descriptions and conclusions about some object; and managing and evaluating the content analysis. For any content analysis effort, the action of delineating information involves identifying the conceptual framework

for the effort; developing plans for obtaining, providing, and applying information; and developing plans for managing and evaluating the effort. Obtaining content analysis information involves the four key tasks of unitizing, sampling, coding, and analyzing messages. Providing the information involves some kind of interaction between the analyst and the audiences for the study. Applying the information depends on the intended uses for the study and any unintended uses that are also attempted. Managing the effort involves allocating and coordinating the personnel, physical resources, time, and money for all the actions of the effort. Evaluating the effort involves characterizing it and establishing its quality. Two key standards of quality for any content analysis effort are the concepts of reliability and validity. The authors of evaluation standards (ERS Standards Committee, 1982; Joint Committee, 1981) consider reliability and validity to be most clearly related to how information is delineated and obtained. Because content analysis is considered to be a quasi-evaluation approach, this relationship should also hold for such studies. As a result, the remainder of this section focuses on delineating and implementing the tasks of unitizing, sampling, coding, and analyzing messages.

Unitizing Messages

Because messages are very complex entities and they are produced in such high volumes, they cannot usually be properly analyzed in their entirety. To address this problem, messages can be divided into three basic types of units: (1) sampling units, (2) coding

units, and (3) context units. "Unitizing involves defining these units, separating them along their boundaries, and identifying them for subsequent analysis" (Krippendorff, 1980, p. 57). Each of these types of units are described next.

Sampling Units. Whenever a class of messages is so large all the messages cannot be included in the analysis, some kind of sample must be drawn. Sampling units should be constructed so that they are independent of each other. "Here 'independent' is synonymous with unrelated, unbounded, unordered, or free so that the inclusion or exclusion of any one sampling unit as a datum in an analysis has neither logical nor empirical implications for choices among other units" (Krippendorff, 1980, p. 57). Three types of sampling units are generally used: (1) sources, (2) complete documents (messages), and (3) sections of documents (Holsti, 1969, P. 130). Sources include people, government agencies, publishers, broadcasters, and other such entities. Complete "documents" include things like books, magazines, reports, interview transcripts, speeches, TV shows, movies, and photographs. Sections of documents include specific pages, editorials, responses to specific questions, time or space segments, and so on. Depending on the size of the complete set of available messages and the design of the analysis, more than one type of sampling unit can be used for any particular study.

Coding Units. Holsti (1969) defines a coding unit as "the specific segment of content that is characterized by placing it in a given category" (p. 116). Most coding units used in content analysis studies can be placed into one of five basic groups: "[1] words,

[2] themes, [3] characters, [4] items, and [5] space-and-time measures" (Berelson, 1952, p. 136).

The word as a coding unit is self-explanatory. It has been used most often in studies on readability, style, psychotherapy, and disputed authorship (Holsti, 1969, p. 116).

The theme is an assertion about an attitude object. "In its most compact form, the theme is a simple sentence, i.e., subject and predicate" (Berelson, 1952, p. 138). Osgood's evaluative assertion analysis (Osgood, 1959; Osgood et al., 1956) is an example of using this type of coding unit. However, because most sentences can be transformed into two or more simple assertions, Holsti (1969) argues, "This process of reducing a grammatical unit into thematic units--sometimes called 'unitizing'--can seriously reduce reliability unless the structural properties of the thematic units are precisely defined" (p. 117). Another name for transforming sentences into simple assertions or themes is "kernelizing" (Krippendorff, 1980, p. 62).

The character as a coding unit usually refers to real or fictional people. More broadly defined as referential units, this category can include "particular objects, events, persons, acts, countries, or ideas to which a particular expression refers" (Krippendorff, 1980, p. 61). This type of unit has most often been used to study the portrayals of different types of characters in different media (Holsti, 1969, p. 117).

An item is "the whole 'natural' unit employed by the producers of symbol material" (Berelson, 1952, p. 141). Examples of items include books, reports, newspaper and magazine articles, speeches,

radio and TV programs, responses to open-ended questions, and so on. Concerning this type of coding unit Berelson (1952) notes, "Analysis by the entire item is appropriate whenever the variations within the item are small or unimportant. . . . But if detailed categories are added which introduce variations within items, the item-unit may become inappropriate" (p. 141).

Like the item, space-and-time measures are examples of physical coding units. In this case, however, items are subdivided by some convenient measuring unit like the column-inch of print, minutes of broadcast time, feet of film or video tape, or grid squares over photographs. Such units have been applied almost exclusively to studies designed to identify the subject matter covered by complete items (Berelson, 1952, p. 143).

Context Units. A context unit is "the largest body of content that may be searched to characterize a recording unit" (Holsti, 1969, p. 118). The size of the context unit can be equal to or larger than the coding unit but never smaller. Perhaps the most common practice is to completely ignore context units (consciously or unconsciously), making it the same size as the coding unit. However, consciously determining the size of the context unit is crucial because its size can affect the results of the analysis. For example, Geller, Kaplan, and Lasswell (1942) demonstrated that as the size of the context unit grew from a single sentence to complete newspaper editorials, the proportion of coding units scored "neutral" diminished. As another example, for the content analysis method called contingency analysis (Osgood, 1959), in which inferences are based on the co-occurrences

of coding units within a larger context unit, the results are highly dependent on the number of words in the context unit. However, Osgood "found contingency values to be roughly constant between 120 and 210 words as units" (p. 62). Krippendorff (1980) underscores the importance of context units as follows:

> Context units set limits to the contextual information that may enter the description of a recording unit. They delineate that portion of the symbolic material that needs to be examined in order to characterize a recording unit. By defining a larger context unit for each recording unit, the researcher recognizes and makes explicit the fact that symbols codetermine their interpretation and that they derive their meaning in part from the immediate environment in which they occur. (p. 59)

Sampling Messages

Although the sampling of messages is usually required in a content analysis effort, the methods available are no different than those for any other social science discipline. Because of this, only one particularly difficult sampling problem for most content analysis efforts is discussed. After an analyst has determined what kinds of materials are most relevant to the study of interest, the volume of such materials probably will still be too large to analyze completely. As a result, sampling of materials is necessary. However, another problem is the material actually available might not truly represent the population of interest for the study. The concern for this issue is widely shared by content analysts (e.g., Berelson, 1952, p. 175; Holsti, 1969, p. 129; Krippendorff, 1980, p. 66). As a result, analysts often must decide if they want to draw a sample

representative of the available materials, or draw a sample representative of an underlying population differentially represented in the available materials. Krippendorff (1980, p. 68) recommends using proportional sampling techniques when making inferences about a population is desired and if it is suspected the available materials do not adequately reflect the population of interest. However, this technique does require at least a tentative hypothesis about how the population is misrepresented by the available materials.

Coding Messages

Holsti (1969) succinctly describes the basic process and role of coding in content analysis as follows:

> Coding is the process whereby raw data are systematically transformed and aggregated into units which permit precise description of relevant content characteristics. The rules by which this transformation is accomplished serve as the operational link between the investigator's data and his [or her] theory and hypotheses. (p. 94)

Because of the complexity of this process, good coding instructions must be explicitly developed to ensure the quality of the effort. Krippendorff (1980) suggests coding instructions should include:

> [a] a prescription of the characteristics of the observers (coder, judges) employed in the coding process; [b] an account of the training these observers undergo to prepare themselves for the task; [c] a definition of the recording [and context] units including procedures for their identification; [d] a delineation of the syntax and the semantics of the data language (variables, categories) including, when necessary, an outline of the cognitive procedures to be employed in placing data into categories; and [e] a description of how data sheets are to be used and administered. (p. 174)

These subtasks and the specifications for their instructions are discussed next.

Observers. Two characteristics of observers (coders) are particularly important if they are to adequately code messages (Krippendorff, 1980, p. 72). First, they must be familiar with the nature of the material to be recorded. This condition is best met when the social background of the coders is similar to the social background of the producers of the material. Second, they must be capable of reliably applying the data language (category system) to the materials. Coders with at least some background in social science methods are desirable, but all coders will need training on the specifics of the coding problem at hand.

Training. Because of the complexity of many content analysis efforts, the training of observers and the development of specifications for coding and context units, data languages, and data sheets is an interactive, iterative process. Although all studies will have their unique features, Krippendorff (1980) presents a set of typical activities for training observers.

> [a] The research designer formulates his [or her] initial data requirements. [b] He [or she] familiarizes himself [or herself] with the way relevant information is expressed in the source material. [c] He [or she] formulates written recording instructions. [d] Working with the coders who are to apply them, instructions are jointly interpreted and modified until they meet suitable reliability requirements. . . . [e] Recording instructions are tested with a fresh set of independent observers. (pp. 73-74)

The final step is intended for studies likely to be replicated by other analysts elsewhere or by the same analysts with a different set

of materials. Without it, the effects on reliability of observers participating in the development of coding instructions cannot be known.

Recording and Context Units. Recording and context units have already been discussed under the task of unitizing. The type of units defined for the study depends on its theoretical framework, hypotheses, and operational transformation into a data language. When the units have easily identifiable physical boundaries (e.g., words, sentences, paragraphs, column-inches, articles, magazines, books, complete responses to questions) the specifications for identifying them are straightforward. However, when the units are primarily symbolic (e.g., themes, assertions, characters, concepts) the instructions must specify what constitutes a unit and the steps needed to identify one. If context units are not the same size as coding units, the steps needed to identify them must also be explicitly specified.

Data Language. The data language constitutes the set of rules and categories through which raw data is operationally linked to the analyst's theories and hypotheses. For the remainder of this paper the term, category system, is used to represent this set of rules and categories. The central role categories play in content analysis is highlighted by Berelson (1952), "Content analysis stands or falls by its categories. Particular studies have been productive to the extent that the categories were clearly formulated and well adapted to the problem and to the content" (p. 147). However, the development of productive categories is by no means a simple task, and

several attempts at identifying the appropriate categories are often required. Holsti (1969) comments on the iterative nature of the process as follows:

> In the absence of standard schemes of classification, the analyst is usually faced with the task of constructing appropriate categories by trial and error methods. This process consists of moving back and forth from theory to data, testing the usefulness of tentative categories, and then modifying them in light of the data. (p. 104)

Krippendorff (1980, pp. 121-123) describes a number of computer-implemented techniques that can be used to help develop useful categories. These techniques are described in the section on performing content analysis tasks with microcomputers.

Adapting Kerlinger's rules of categorization (e.g., 1986, pp. 127-130), Holsti (1969) further specifies five criteria good categories should meet: "categories should [1] reflect the purposes of the research, [2] be exhaustive, [3] be mutually exclusive, [4] independent, and [5] be derived from a single classification principle" (p. 95).

Categories must first and foremost reflect the purpose of the research, otherwise, coding the data has no meaning. Holsti (1969) elaborates as follows:

> This means, first of all, that the analyst must define clearly the variables he [or she] is dealing with (the "conceptual definitions"), and secondly, he [or she] must specify the indicators which determine whether a given content datum falls within the category (the "operational definition"). A good operational definition satisfies two requirements: it is a valid representation of the analyst's concepts, and it is sufficiently precise that it guides coders to produce reliable judgments. (p. 95)

The requirement that categories be exhaustive means all coding units must be capable of being placed into a category. If this were not the case, coding units would be systematically excluded from the analysis--biasing the results.

The condition of mutually exclusive categories requires each coding unit to be placed into no more than one category. This means the operational definitions of variables must be unambiguous, and the amount of information in a coding unit must approximate the amount of information required to assign it to a single category.

The requirement for independence of assigning coding units to categories reflects one of the basic assumptions of many statistical analysis techniques. When complete messages are divided into many coding units, this assumption might be violated. In such cases, to the extent subsequent analyses rely on the assumption of independently assigned coding units, the results will be biased.

The criterion that categories should be derived from a single classification principle provides the basis for constructing mutually exclusive categories. When a group of categories reflects two or more underlying variables as if they were one, individual coding units can usually be assigned to more than one category.

For example, the four categories, positive, negative, teachers, and superintendent, reflect two variables--evaluation and staff position. Because the statement, "The teachers work their tails off every day," applies to both the positive and teachers categories, the category system violates the criterion of mutually exclusive categories. The common remedy is to "cross" the two variables (e.g.,

Kerlinger, 1986, pp. 129-130; Holsti, 1969, pp. 100-101) so that the set of categories reflects all combinations of the possible values for each variable. In the above example, the categories would become positive-teachers, positive-superintendent, negative-teachers, and negative-superintendent.

An alternative approach is to use a branching decision scheme. "Decision schemes regard each datum as the outcome of a predefined series of decisions" (Krippendorff, 1980, p. 77). He lists the advantages of decision schemes as follows:

> First, decision schemes can avoid problems arising from categories that are on different levels of generality or overlapping in meaning. . . . Second, when recording units are multidimensional, decision schemes offer the opportunity of decomposing a complex judgment into several simple decisions and thereby achieve levels of reliability not obtainable otherwise. Third, decision schemes drastically reduce the number of alternatives to be simultaneously considered at each step. (pp. 77-78)

For the above variables, evaluation and staff position, if we are interested in focusing in on the targets of negative evaluations, but we do not care about the targets of positive evaluations, we could use a two-step decision scheme as follows: (1) divide the evaluative statements into positive and negative groups, then (2) divide the negative statements into those about teachers and those about the superintendent. This decision scheme requires only three categories (positive, negative-teachers, and negative-superintendent) and it meets the criterion of mutually exclusive categories. Thus the statement, "The teachers work their tails off every day," would be unambiguously placed into the positive category. Examples of many

other kinds of category systems are summarized in a number of sources (e.g., Berelson, 1952, pp. 147-168; Dunphy, 1966; Holsti, 1969, pp. 95-116; Krippendorff, 1980, pp. 75-81, 85-118).

Data Sheets. Data sheets are the medium on which the coding process is documented. Because of this, they must contain all the information necessary to represent the placement of coding units into categories, as well as some "housekeeping" information. Krippendorff (1980) suggests data sheets should contain at least three types of information: "[1] administrative information, [2] information on data organization, and [3] information on the phenomena to be recorded, the data" (p. 82). His examples of administrative information include: (a) identification of the project to which the data apply, (b) identification of the stage of processing of the data, (c) identification of the individuals who processed the data at each stage, and (d) instructions on how the data should be transferred to computer data files. Information on data organization refers to how data on a particular sheet relates to data on other sheets. For example, demographic information on data sources is usually kept on one type of sheet and the messages they produce are usually recorded on another type of sheet. Unique identification numbers for the sources placed on both sheets maintain the relation between the sources and their messages. Finally, the data (coding units) are on the sheets. Depending on the size of the context units, they might also be on the same sheets. The representation of the data and possible categories into which they can be coded depends on the exact nature of the data and the category system used. Because references

to examples of different category systems have already been provided, they are not repeated here.

Analyzing Messages

As reflected in the first criterion for constructing good categories, the nature of the analysis of a set of messages should be intimately related to the purposes of the effort. The previous discussion on the uses of content analysis summarized three general purposes for conducting a content analysis. They were to: (1) make inferences about the antecedents of communications, (2) describe and make inferences about the characteristics of communications, and (3) make inferences about the consequences of communications. Holsti (1969, pp. 27-37) presents a series of general designs that can be use to organize content analysis efforts directed toward each of these purposes. The designs of actual content analysis studies should reflect their general purpose and the specific theories, hypotheses, and variables that apply. Because of the interdependent nature of designing category systems and analytical procedures, the references listed for examples of category systems also contain many examples of specific analytical procedures.

Performing Content Analysis Tasks with Microcomputers

Microcomputer programs can be of use for many kinds of administrative tasks. Because of this, most research and evaluation organizations are likely to already have a battery of microcomputers and general purpose programs available to them. Gray (1984d) lists

five types of "administrative" or management programs of use to the field of evaluation: "[1] word processing, [2] calc/statistics, [3] data base management, [4] communications and networking, and [5] graphics" (p. 80). Two types of programs--word processing and data base management--are particularly relevant to this discussion. The reasons for this are discussed later in this section. Elsewhere, Gray presents more detailed discussions of how word processing programs (1984a) and data base management programs (1984b, 1984c) can be used for administrative purposes in the field of evaluation.

These general uses for microcomputer programs are important because they provide the basic justification for purchasing the microcomputers and programs in the first place. However, once they are available to the organization, these resources can also be adapted to content analysis uses.

Besides general project management and support activities like word processing and data base management, computers can be put to three general uses in content analysis efforts (Krippendorff, 1980). They include statistical analyses, computational aids for survey and discovery, and computational content analysis (pp. 119-128).

Statistical analyses are not unique to content analysis efforts and they are not of particular interest here. Common descriptive and inferential statistics familiar to social scientists in general are also of use in many content analysis studies.

Computational aids for survey and discovery help content analysts consolidate large masses of textual material so that various types of overviews of the information contained in them can be

developed. Such overviews can be used during the development of coding instructions to help identify applicable categories. Similar techniques can also be used to help place sets of coding units into existing categories. In computational survey and discovery, the human still makes all the "hard decisions" and simply uses the computer to perform a number of "clerical" functions. This is the use of computers of interest for this study.

Computational content analyses are performed primarily by computer programs rather than by humans. Such programs are simultaneously very complex and overly simplistic. That is, the programs themselves are very large and complicated, requiring high powered mainframe or supermini computers; while their performance is usually narrowly focused and often lacking the "common sense" of even a novice content analyst. The best example of this high powered type of program is the General Inquirer (Stone et al., 1966).

While Krippendorff and other authors (e.g., Gerbner, Holsti, Krippendorff, Paisley, & Stone, 1969; Holsti, 1969, pp. 150-194) discuss a number of variations of computer-assisted content analysis, the techniques most useful for survey and discovery can be placed into three basic groups: (1) key words out of context, (2) key words in context, and (3) information retrieval. All these techniques can be implemented with custom-designed computer programs running on large or small computers. However, some of them can also be implemented on programs originally designed for other purposes. Variations of each technique and some of the general purpose programs that can be used to implement them are discussed next.

Key words out of context are basically word lists. The lists are usually of single words but they can also be of phrases or groups of words that occur within a specified distance of each other (e.g., no more than five words apart). The frequency of occurrence of each item in the document is also listed. The list can be ordered alphabetically or by frequency of occurrence. Finally, items with high, low, or chance frequencies of occurrence; or types of words like articles, prepositions, and pronouns; can be deleted from the list.

Word lists are relatively easy to produce for a skilled computer programmer with just about any programming language, such as BASIC or Pascal. However, because of the way some "spelling checker" programs are designed, they automatically produce word lists. If these lists are accessible to the user, they can also be considered key word out of context lists. One such program for microcomputers is called The Word Plus (Holder, 1982). An option of this program is to create a text file that lists all the unique words contained in a different text file (p. 38). The number of times each word appears in the source file (e.g., a collection of coding items) is also included in the list. The list can be ordered alphabetically or by frequency of occurrence. Because this list is a text file, it can be edited with a word processing program. This means unwanted words like articles, pronouns, or those with low frequencies can be easily removed from the list. The list can then be used to help decide what categories should be used in the final content analysis.

Key words in context are lists of specified words surrounded by parts of the text in which they occur. This shows the reader how the

word was used in context. The length of the text is usually short enough to be printed on one line with the key word centered. The line can also be indexed so the source material is easily accessible. This is a very special type of list that is more difficult for a programmer to produce than simple word lists. In addition, no computer programs designed for general business or educational uses produce this kind of list. Wood (1984) summarizes several qualitative and quantitative social research uses for this type of list.

The third type of technique, information retrieval, can be used on "original" documents, such as complete word processing files, or textual data base files in which each "record" can contain one coding unit identified from a larger document. The two most common information retrieval functions that can be performed on these files are searching for and sorting information. Once found or sorted, the information can then be displayed to the user in any number of ways.

When the material is basically "free form," like the chapter of a book, information retrieval is primarily limited to searching for and displaying specified words or phrases. Just about any word processing program has this capability, although the results usually can only be presented on the screen. Depending on the particular word processing program, a section of text containing the specified items could be "cut" from the document and then "pasted" into a different document with similar passages, but the following approach is much more powerful and convenient overall.

When the text (recording unit) is organized into a database as one "field" within a larger "record," both searching and sorting can

take place on any one or a combination of fields. This allows for very flexible and powerful manipulations of the textual material with relatively little effort on the part of the user, particularly when the other fields of a record contain relevant information about the textual material. A further advantage of this approach is the results of searches and sorts can be sent to a number of destinations, such as the screen, printers, and other files. Examples of both of these capabilities in a microcomputer word processing program are WPS List Processing (Digital Equipment Corporation, 1984a) and WPS Sort (Digital Equipment Corporation, 1984b). An example of a microcomputer data base management program with searching, sorting, and displaying capabilities is dBASE II (Ratliff, 1982). This technique can be used during the process of developing a category system or while coding units for an existing set of categories.

Many evaluation and research oriented organizations now have microcomputers with a number of general purpose, business application programs like those for word processing and data base management. As a result, they also already have a basic library of programs that can be adapted to many survey and discovery uses in content analysis efforts. The knowledge of a few simple techniques and a lot of imagination are the keys to discovering these uses.

Summary

This chapter was used to present the conceptual and operational relationships between evaluation, content analysis, and microcomputers. The discussion was divided into four major topics: (1) the

conceptual relationships between evaluation and content analysis, (2) an overview of evaluation approaches, (3) a general model for conducting an evaluation effort, and (4) an overview of content analysis.

The key concepts for identifying the relationships between evaluation and content analysis were organized into three basic groups: (1) information, (2) actions, and (3) standards of quality. The relationships between these concepts and the two fields were summarized by presenting working definitions of evaluation and content analysis that used the concepts and conformed to a common grammatical structure. Both definitions are repeated here. Good evaluation is the high quality process of delineating, obtaining, providing, and applying characterizations and appraisals about some object; and managing and evaluating the evaluation. Good content analysis is the high quality process of delineating, obtaining, providing, and applying descriptions and conclusions about some object; and managing and evaluating the content analysis. These definitions contain direct references to the information and action components that were identified. However, because of the emerging status of standards of quality in both fields, only indirect references were made to them in the definitions. In any event, the two standards of quality currently found in common between the two fields are those of reliability and validity.

The overview of evaluation approaches was organized around a classification framework with three dimensions: (1) epistemology, (2) major perspective, and (3) orientation to values. Although this

framework contained twelve cells, fifteen general evaluation approaches were found to fill only seven of them. Within this framework, content analysis was found to be an objectivist, elite, quasi-evaluation approach. Each of the evaluation approaches were further summarized in terms of their main organizer, purpose, strengths, and weaknesses.

The general model for conducting an evaluation effort provided a graphic representation of the working definition of evaluation. As such, it represented the key components and their logical relationships for an evaluation effort. A decision network based on the general model was also presented. It represented the dynamic and iterative nature of processing evaluation information. Because content analysis was classified as a quasi-evaluation approach, this model also applies to content analysis efforts.

The overview of content analysis contained four main discussions: (1) the classification of content analysis uses, (2) a summary of content analysis uses, (3) a presentation of the tasks crucial to delineating and obtaining high quality content analysis information, and (4) the ways microcomputers can be used to help delineate and obtain content analysis information. Fifteen content analysis uses were classified in relation to the basic components of the communication paradigm. This paradigm can be represented by the compound question, "Who, said what, to whom, why, how, and with what effect?" In the summary, groups of content analysis uses were discussed in terms of their general purposes and the communication question they addressed. Specific uses were discussed to the extent

they could be related to the field of evaluation. One such use was to make inferences about the attitudes of a selected group of people based on the messages they produced. A common way to collect such messages in evaluation efforts is to ask them to respond to a set of open-ended survey questions.

Four key tasks involved in obtaining content analysis information for any purpose were further described. These tasks included unitizing, sampling, coding, and analyzing messages. Finally, three ways of using computers in content analysis tasks were identified. One of the ways, as computational aids for survey and discovery, was considered to be the most promising avenue for using microcomputers to help obtain content analysis information, particularly when developing coding instructions or actually coding messages. Two survey and discovery techniques that can be implemented using certain kinds of microcomputer programs not specifically designed for content analysis were identified. They included: (1) producing key words out of context lists with certain kinds of spelling checker programs, and (2) performing information retrieval activities such as searching for, sorting, and displaying messages with certain kinds of word processing programs or data base management programs.